

Use of Monocular Groupings and Occlusion Analysis in a Hierarchical Stereo System*

RONALD CHUNG

Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong

AND

RAMAKANT NEVATIA

Institute for Robotics and Intelligent Systems, University of Southern California, Los Angeles, California 90089-0273

Received April 15, 1992; accepted August 18, 1994

We describe a hierarchical stereo system that computes a hierarchy of descriptions up the surface level from each view using a perceptual grouping technique and matches these features at the different levels. With the description and correspondence processes going hand in hand, we allow high-level abstract features to help reduce correspondence ambiguities, and we exploit the multiple views to help confirm the different levels of descriptions in the scene. Occlusion is a major problem in stereo analysis and is often not treated explicitly. We present a basic property of occlusions in stereo views and show how we can use this property incorporated with the structural descriptions to identify different types of occlusions in the scene. In particular, we identify depth discontinuities and limb boundaries and infer properties of surfaces that are visible only in one of the stereo views. We give some experimental results on scenes with curved objects and with multiple occlusions. © 1995

Academic Press, Inc.

1. INTRODUCTION

A central problem of computer vision is to recover 3-D descriptions of objects in a scene from 2-D images. As depth information along the line of sight is lost during the projection of a point in 3-D onto a 2-D image, the recovery problem from a single 2-D image is underconstrained. Stereo vision, in which multiple views of a scene are used, has become a common method for reconstructing 3-D in-

formation. In this paper we restrict our discussions to binocular vision which requires only two views of the scene. The essence of stereo vision is that once a feature in one view and another feature in the other view are known to be projections of the same physical point in the scene, the intersection of the two lines of sight through the features will determine the 3-D position of the physical point.

While stereo provides one possible solution to depth determination from 2-D images, the existence of multiple views brings another problem. For each feature in one view, we must locate the feature in the other view which is the projection of the same physical entity in 3-D. This is known as *the correspondence problem* and is historically regarded as the most important problem in stereo analysis.

The ultimate goal of vision, however, is not just to reconstruct a set of uncorrelated depth estimates, but to come up with some meaningful, symbolic descriptions of the imaged scene so that we can represent, store, match, and retrieve them easily for tasks such as object recognition and grasping. Such descriptions should distinguish objects in the scene from the background (*segmentation*) and represent the interaction of the surfaces that compose an object (*shape decomposition*). What is important in 3-D is whether we can extract the *depth and orientation discontinuities* among the different surface patches, and whether we can identify the nature of the visible surface boundaries as *crease boundaries* or *limb boundaries* (boundaries projected from curved surfaces slowly turning away from the viewer). We will call this *the description problem*.

While much work has been done on the correspondence problem to reconstruct depth estimates, few have addressed the description problem in stereo analysis. Most take the view that stereo correspondence *precedes* structural descriptions of the scene (other than in extracting simple low-level features such as edge elements) and that

* This research was performed at the Institute for Robotics and Intelligent Systems, University of Southern California. It was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract F49620-90-C-0078. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

the two processes are more or less independent. We take the view that while for some scenes such as random dot stereograms, the correspondence process must come before the description process, in general structural descriptions even in the monocular stage can aid in the correspondence process, and there are many advantages to the two processes working together cooperatively. We propose a hierarchical system that computes a hierarchy of descriptions such as edges, curves, symmetries, and ribbons from each view and matches these features at the different levels. With the description and correspondence processes going hand in hand, we gain the following advantages:

1. high-level abstract features help reduce ambiguity of multiple matches in the correspondence process;
2. correspondences help confirm the different levels of abstract features in the two views and maintain a consistent descriptions about the scene;
3. output is a set of segmented surfaces.

One of the difficulties in pursuing this approach is that it requires the ability to make structural descriptions from monocular views, a traditionally difficult problem. We believe that even though the monocular description problem remains a difficult one, enough progress has been made to give useful descriptions for aiding in the stereo process; we use a method based on the perceptual grouping technique developed by Mohan and Nevatia [24, 25].

Yet merely going to higher level features does not solve the entire problem. *Occlusions* still present difficulties to stereo analysis. Without occlusions in the scene, the stereo correspondence problem can be cast as a one-to-one mapping problem between two sets of features, in which the topology among the features is always preserved if noise is omitted. The traditional use of the stereo constraints of surface continuity and ordering (the surface-continuity constraint assumes that disparity is continuous; the ordering constraint requires that the order of features along epipolar lines be preserved) is in fact valid only in the absence of occlusions. However, if there are multiple objects occluding one another in the scene, some features in the scene may be visible in one view but hidden in the other, and they destroy the invariance of the topology among the features even in the absence of noise. Moreover, a major consequence of occlusion in stereo views is that a closer surface hides a more distant surface to different extents in the two views. This implies that even if two patches in the two views are in fact projected from the same surface, their apparent boundaries do not fully correspond if the physical surface in space is occluded. As a result in addition to looking at the properties of the high-level features alone, we need to study how occlusions affect the interaction among the features to match them across stereo views.

Despite these observations, occlusions are seldom treated explicitly and it is usually assumed that they can be identified either by setting some thresholds during the surface interpolation step or by looking at the disparity differences of the neighboring edge elements [20, 29]. However, even if we can assume the initial depth estimates are perfect, surface interpolation without knowing which side of an edge is the occluded surface would lead to erroneous result. Moreover, the sparsity of features would present difficulties in adjusting the thresholds to detect discontinuities (points on a 1-D step function will be more "collinear" when they are farther apart). It is our belief that occlusions deserve much more attention and must be treated explicitly, not only because they are the major cue that makes the stereo correspondence problem a much harder mapping problem, but also because depth discontinuities that come with occlusions capture some of the most important information in the depth map of a scene. As we shall see in later sections, even the problem of limb boundaries [19] can be cast as a problem of occlusions, and curved surfaces can be identified by examining some effects of occlusions in stereo.

In this paper we describe how perceptual grouping can be used to extract features up to ribbon level monocularly from the stereo views, as proposed by Mohan and Nevatia [24, 25], and what the physical constraints are to match the features hierarchically. Specifically, our system applies the surface-continuity and ordering constraints within a surface only, and not across the occlusion boundaries. We also describe a basic property of occlusions in stereo views, and how we can use this property to identify different types of occlusions. In particular, we use the property to identify depth discontinuities and limb boundaries, as well as to infer properties of surfaces that are visible in only one of the stereo views.

The organization of the paper is as follows. In Section 2 we present some previous work in stereo analysis. In Section 3 we describe a basic property of occlusions in stereo views, and how different types of occlusions can be identified using this property. Section 4 describes our hierarchical stereo system which consists of three subsystems, namely the monocular groupings, the binocular stereo correspondence, and the feedback from stereo correspondence to monocular groupings. Finally, some experimental results are given in Section 5. An earlier shorter version of the paper has been presented in [5].

2. PREVIOUS WORK

Stereo correspondence is complicated by three inherent problems:

1. ambiguity among false targets;
2. photometric distortions and camera noises; and

3. occurrence of occlusions among surfaces.

To deal with this correspondence problem, features of various abstraction in combination with various sets of constraints have been proposed for matching. A good survey for recent literatures can be found in [7]. Two principal approaches are generally used: area-based matching or feature-based matching.

Area-based matching attempts to match small windows from the left and right views by correlating their intensities or the derivatives of their intensities. Representative examples are systems of [6, 14]. To reduce correspondence ambiguity, a coarse-to-fine strategy is generally employed in the matching, in which correspondence results at lower resolutions are propagated to processings at higher resolutions. Area-based matching has the advantages that the intensity windows being matched are simple to extract, and it delivers a relatively dense depth map if the scene is densely textured enough. On the other hand, it requires presence of significant texture in the scene. As information being matched corresponds directly to intensity variations, area-based matching is sensitive to photometric distortions and camera noises. The presence of occluding boundaries in the correlation window will also confuse the correlation-based matcher, giving an erroneous depth estimate. Regarding the coarse-to-fine strategy in stereo matching, Prazdny [27] also argued that high spatial frequencies and low spatial frequencies may be sometimes orthogonal to each other (imagine one seeing through the narrow slits of a fence to a tree far away).

Feature-based matching uses structural features rather than image intensities for stereo matching, and hence is more stable toward photometric variations and capable of handling nontextured scenes. It is also known that edges can be estimated to subpixel accuracy. Feature-based matching can be further subclassified according to how abstract the features being matched are.

Low-level feature-based methods [9, 12, 21, 22] generally match zero-crossings or other types of edge elements. As low-level features contain relatively little information to resolve correspondence ambiguity, use of higher level features has become increasingly popular.

Some [1, 23] have suggested using segments for stereo matching. Ganapathy [10] has implemented a system for matching junctions of polyhedral objects. Systems have also been developed for specific domains such as architectural scenes, and task-specific knowledge is employed for stereo correspondence. For example, Baker *et al.* [2] and Liebes [17] have made use of orthogonal trihedral vertices, whereas Herman and Kanade [13] have made use of vanishing points of horizontal and vertical lines in the scene.

In an important work, Lim and Binford [18] were the first to advocate the use of a hierarchy of features, including high-level features such as surfaces. They also studied the

effects of errors caused by limb boundaries and suggested solutions that are valid under some assumptions. However, this system was demonstrated with only relatively simple scenes. We believe that its major weakness was in the modules to infer surface level descriptions from monocular images. Mohan and Nevatia [25] described a system to extract good surface level descriptions from monocular views of complex scenes, by using a perceptual grouping technique. They also showed results on how such descriptions could be used for stereo matching. Yet no coupling between monocular groupings and stereo correspondence have been suggested.

Little and Gillett [20] were among the first to address explicitly the notion of occlusion in stereo analysis, and they came up with a method to locate depth discontinuities from stereo. The method decides whether an edge element is on a depth discontinuity or not merely from the disparity information of the neighboring edge elements, and therefore requires significant texture to be present both in the foreground and the background around the edge. After the preliminary record of our work appeared [5], there were a few other work on occlusion analysis [3, 11]. However, like the method of Little and Gillett, all of them require the scenes to be significantly textured as decisions are made based upon local disparity information. Dhond and Aggarwal [8] have also addressed the specific notion of narrow occluding objects in stereo analysis.

To conclude, previous work generally suffer from three inadequacies:

1. Although high-level features such as segments and curves have been proposed for stereo matching, they are not high enough to reduce all correspondence ambiguity, especially when there are repetitive patterns in the images.
2. Constraints in stereo correspondence such as the surface-continuity and ordering constraints are in fact applicable only for features within a surface, but they are usually applied over the entire scene, even across occlusion boundaries.
3. The output is usually merely a set of depth estimates, and the description problem is never addressed.

For these reasons we agree with [18, 25] that hierarchical features up to surface level should be used for stereo matching. As surfaces may have markings and their boundaries may be fragmented and occasionally missing, extracting surfaces by merely doing contour-tracing would not be reliable. Mohan and Nevatia [24, 25] have suggested using a technique called perceptual grouping to extract surfaces from an image, in which relationships of cocurvilinear and symmetry are employed to extract hierarchical features up to the ribbon level.

We use the perceptual grouping technique proposed by Mohan and Nevatia, though we have made important addi-

tions to the grouping technique itself. We use a hierarchy of descriptions as advocated by Lim and Binford, but we believe that we have developed a more competent system, in part because of the better descriptions we are able to generate and in part because of a more sophisticated occlusion analysis that we employ. Our matching technique is also quite different from that of Lim and Binford.

We impose the following restrictions for the scene:

1. all surfaces are opaque;
2. every vertex is formed by not more than three surfaces;
3. the images are taken under general points of view; i.e., in each view the observed structures such as surfaces and junctions are stable under slight perturbation of the viewpoint;
4. the resolution of the images is high enough compared with the sizes of the objects in the scene such that errors in detecting edges around junctions are negligible.

3. OCCLUSION EFFECTS IN STEREO

We live in a 3-D world which is full of opaque objects and surfaces. From any point of view in space, it is inevitable that closer surfaces will occlude or hide more distant surfaces to various extents. Occlusion is therefore a fundamental phenomenon that must be overcome by any stereo system.

The general conception about occlusion is one with a depth discontinuity as the occlusion boundary. Here we take the view that the scope of occlusion can be expanded without restricting the nature of the occlusion boundary, as long as it is within the scope of having one surface patch blocking the sight of another surface patch. Further, we categorize occlusion into different types according to whether the occluding and the occluded surface patches are detached, and if they are connected, whether it is smooth (differentiable) across the boundary between them.

The reason for the generalization is that, regardless of the nature of the occlusion boundary, as long as there is a surface patch blocking the sight of another surface patch, all types of occlusion share the same property in stereo views. This property is the major source of problems in stereo correspondence, and in turn we can make use of it to recover the nature of the occlusion boundary. This will be explained in detail later in this section.

We term the different types of occlusion as the following:

1. *Depth Discontinuity Occlusion.* It is an occlusion in which the occluding and the occluded surface patches are detached. The occlusion boundary is therefore a depth discontinuity. Accompanying the occlusion are *T junctions* along the depth discontinuity (Fig. 1a).

2. *Orientation Discontinuity Occlusion.* It is an occlu-

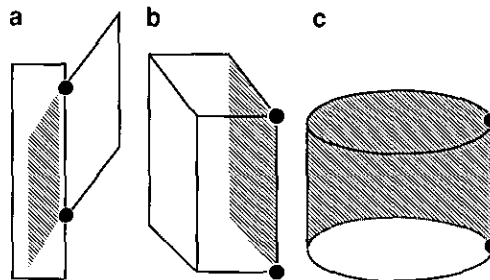


FIG. 1. Types of occlusions: (a) depth discontinuity occlusion (with *T junctions*); (b) orientation discontinuity occlusion (with *W* or *L junctions*); (c) limb occlusion (with limb-*W* or limb-*L junctions*). (Shaded areas are occluded regions.)

sion in which the occluding and the occluded surface patches are connected, but it is not differentiable (the boundary is a fold) across the boundary between the surface patches. The occlusion boundary is therefore an orientation discontinuity. Accompanying the occlusion are viewpoint-independent junctions such as *L*, *W* (Arrow), or *Y* junctions along the orientation discontinuity (Fig. 1b). We call these junctions *edge junctions*, as they are formed from real edges.

3. *Limb Occlusion.* It is an occlusion in which the occluding and the occluded surface patches are connected, and it is also smooth across the boundary between the surface patches. The occlusion boundary is therefore a limb boundary. Accompanying the occlusion are limb-*W* or limb-*L junctions* along the limb edge, provided that the limb terminates at terminator surfaces (Fig. 1c). We call these junctions *limb junctions*, as they are formed on a limb.

Note that more than one type of occlusion can occur simultaneously along the same edge in an image; e.g., a surface can occlude an adjacent surface along an orientation discontinuity and another surface along a depth discontinuity on the same edge. We define an occlusion as one occurring between two specific surfaces along a specific continuous occlusion boundary, and multiple occlusions may have their occlusion boundaries projecting to the same edge.

3.1. A Basic Property of Occlusion in Stereo Views

As mentioned above, there is a common property shared by all types of occlusion in stereo views. Regardless of the nature of the occlusion boundary, at the vicinity of the occlusion boundary there is a region in one view which does not have correspondence in the other view. Such a region is in fact the projection of a portion of the occluded surface patch which is visible only in one of the stereo views. We call such a region the *not-visible-from-other-view (NVFOV) region*.

An explanation for the occurrence of the NVFOV re-

gions is that, as the angle of view moves from one view to the other, different surfaces at different depths will move in their images to different extents, thereby generating such NVFOV regions on the stereo images (the formation of NVFOV regions have close resemblance with the formation of shadows, but we do not call them shadow regions to make them distinct from real shadows). With respect to each occlusion boundary between any two surfaces, we will call the view containing the NVFOV region the *more-exposure view*, as it shows more of the occluded surface, and the other view the *less-exposure view*. We will also call the junctions along the occlusion boundary in the more-exposure view the *more-exposure junctions*, and the corresponding junctions in the other view the *less-exposure junctions*.

It is the occurrence of these NVFOV regions, rather than occlusions themselves, that presents difficulties to stereo correspondence. In two ways they hinder stereo correspondence (Fig. 2):

- *Intersurface Disparity Discontinuity.* The NVFOV regions disturb the adjacency of points across occlusion boundaries and destroy the surface continuity assumption commonly used in stereo correspondence methods. In fact it is these NVFOV regions which create disparity discontinuities across different surfaces.

- *Intrasurface Stereo Noncorrespondence.* A closer surface generally hides a more distant surface to different extents in different views. This means that even if two regions in the two views like $ABCD$ and $E'F'C'D'$ in Fig. 2a are truly projected from the same physical surface in 3-D, their apparent boundaries and also points within the regions may not be all corresponding if the physical surface is occluded. The points that do not have correspondence are those that are bounding and within the NVFOV region $E'F'B'A'$ caused by the occlusion.

These NVFOV regions can occur with all three types of occlusions, and their significance determines directly how

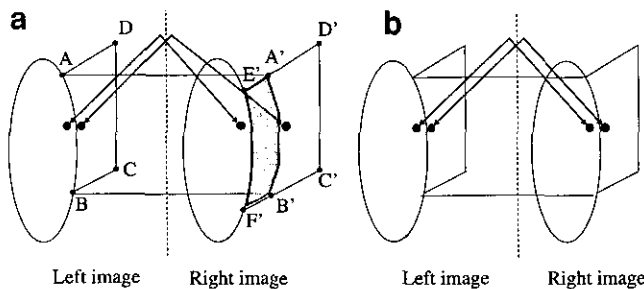


FIG. 2. Effects of NVFOV regions to stereo correspondence. (a) Significant NVFOV region: disparity is discontinuous across occlusion boundary and images of occluded surface are not totally corresponding; (b) negligible NVFOV region: all problems due to occlusion disappear.

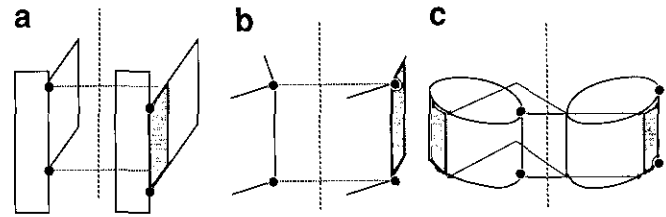


FIG. 3. Effects of NVFOV regions on the junctions along occlusion boundaries: (a) at depth discontinuity occlusion, T junctions are displaced across epipolar lines; (b) at orientation discontinuity occlusion, junction types change (between W and Y, and between L and W in this figure); (c) at limb occlusion, limb junctions are displaced across epipolar lines. (Shaded areas are the NVFOV regions.)

much interference occlusions will present to stereo correspondence (Fig. 3). From this perspective, problems in stereo vision such as inapplicability of surface-continuity constraint across occlusion boundaries, noncorrespondence of limb edges, and occurrence of NVFOV surfaces next to orientation discontinuities are merely different aspects of the occlusion problem. Although occlusion is such an important issue in stereo vision, little attempt has been made to handle it explicitly in previous stereo systems. Two of the previous systems [18, 25] do extract surfaces and then T junctions from each view so that real physical boundaries of the surfaces can be identified for stereo matching. This handles depth discontinuity occlusions but *not limb occlusions and orientation discontinuity occlusions*.

An important observation is, as occlusions generate an NVFOV region in stereo views, the occurrence of the NVFOV region in turn modifies the junctions accompanying the occlusion (Fig. 3). More precisely, it moves the viewpoint-dependent junctions (T junctions and limb junctions) across the epipolar lines and changes the junction types of the viewpoint-independent junctions (edge junctions). The behaviors of different junctions in stereo views can be used as means to identify different types of occlusions. In the following we will present in detail how the NVFOV region modifies the junctions and how the occlusion types can be identified.

3.2. A Constraint for the NVFOV Region

If an NVFOV region is to appear along with an occlusion, in which view is it going to appear? It is mentioned that the occurrence of an NVFOV region creates disparity discontinuity between the occluded and occluding surfaces along the occlusion boundary. Whether around the occlusion boundary the disparities of the occluded surface are larger or smaller than those of the occluding surface will depend upon two factors (Fig. 4): (1) in which direction the NVFOV region shifts the image of the occluded surface relative to the image of the occluding surface, to the left

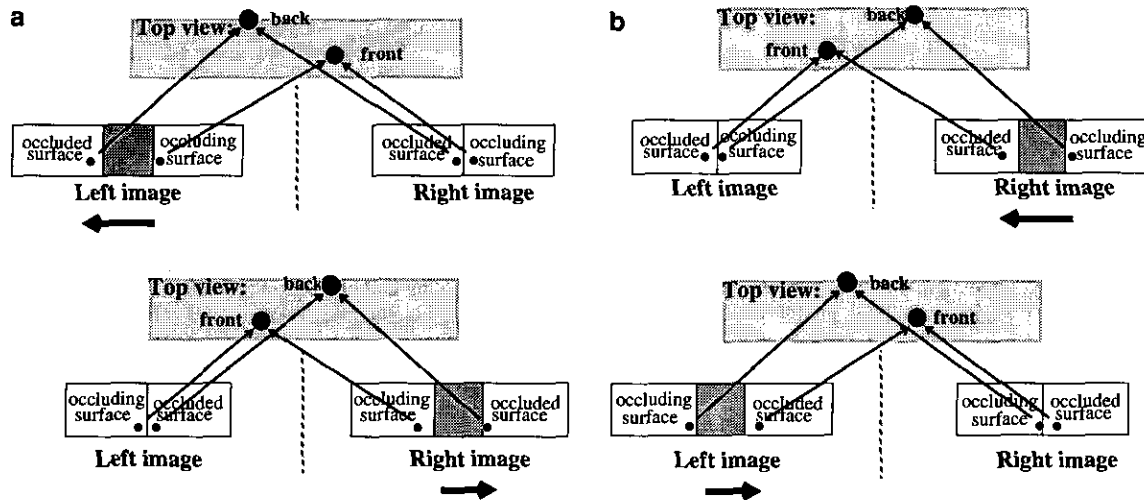


FIG. 4. Asymmetry of the occurrence of NVFOV region with respect to stereo views: (a) possible NVFOV regions; (b) impossible NVFOV regions. (\rightarrow : direction in which image of the occluded surface is shifted relative to that of the occluding surface.)

or to the right; (2) in which view the image of the occluded surface is shifted, the left view or the right view, which is effectively the view where the NVFOV region appears.

Note that the only place where the NVFOV region appears is between the occluded and occluding surfaces in one of the stereo views. This means if the occluded surface is toward the left of the occluding surface, the NVFOV region will always shift the image of the occluded surface to the left relative to the image of the occluding surface. It will be to the right otherwise. Therefore, it can be concluded that given the position of the occluded surface relative to the occluding surface, whether the NVFOV region appears in the left view or the right view directly determines whether the disparities of the occluded surface are larger or smaller than those of the occluding surface around the occlusion boundary.

The fact is, at the vicinity of the occlusion boundary, the occluded surface must be more distant than the occluding surface. This renders the occurrence of the NVFOV region asymmetric with respect to the stereo views, which can be formulated as:

NVFOV Region Occurrence Constraint. The NVFOV region generated by an occlusion in stereo views must appear in the left view if the occluded surface is toward the left of the occluding surface, and in the right view if the occluded surface is toward the right of the occluding surface.

In other words, once we know the relative positions of the occluding and the occluded surfaces, we can easily tell whether the left view or the right view will be the more-exposure view. Likewise, once we know which branches of a junction are from the occluding surface and from the occluded surface, we can tell whether the junction in a

specific view is a more-exposure junction or a less-exposure junction by looking at the relative positions of the branches. We will elaborate on this further for different types of junctions as we mention how NVFOV regions modify them.

3.3. Behavior of Junctions in Stereo Views

Here we present, for all three types of junctions, how we can tell whether a certain junction in a certain view (left or right) is a less-exposure junction or a more-exposure junction with respect to the occlusion, and how the NVFOV region affects the behavior of the junction in stereo views. Note that under the *general viewpoint assumption* in each image (as stated in Section 2), the contours in 3-D involved in the formation of a viewpoint-dependent junction (the limb junction and the T junction) must be differentiable (smooth) at the point that projects to the junction. If not, the junction is considered formed at an accidental viewpoint and will not be considered.

3.3.1. T Junctions

A T junction is formed along an occlusion boundary where the occluding and the occluded surfaces are not connected, as shown in Fig. 5. It is composed of two branches, the stem and the cap, which are boundaries of the occluded and occluding surfaces respectively. As the stem and the cap can be easily identified, we can tell whether a T junction in a specific view is a less-exposure junction or a more-exposure junction by checking whether the stem is toward the left or toward the right of the cap (Fig. 6).

As the viewpoint moves from the position of the less-exposure view to that of the more-exposure view, the

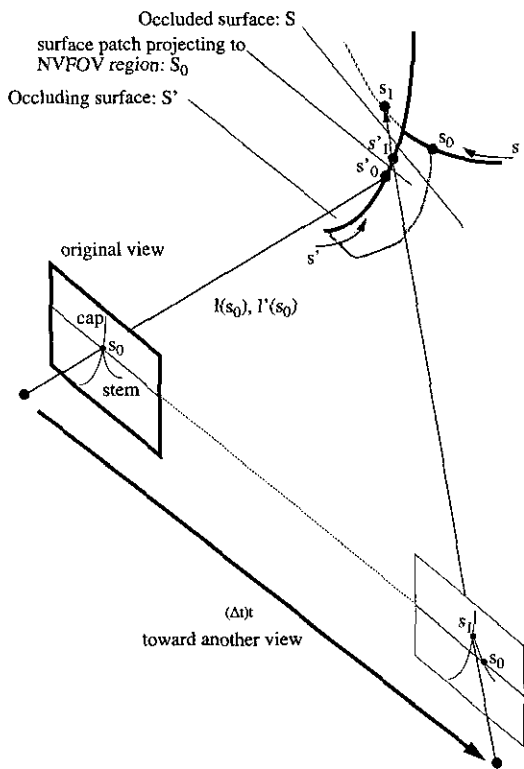


FIG. 5. Behavior of a T junction upon change of the viewpoint.

NVFOV region gradually appears and moves the T junction along the stem branch toward the occluded end. As a result, the T junction will go up or go down across the epipolar lines at the less-exposure view according to the orientation of the vector \mathbf{v} , where \mathbf{v} is the tangential vector to the stem edge at the less-exposure junction and in the direction toward the occluded end (Fig. 7).

We thus have the following constraint:

THEOREM 1. T-JUNCTION EPIPOLAR DISPLACEMENT CONSTRAINT. *A T junction generally moves across the epipolar lines as the viewpoint changes, and the immediate epipolar displacement as the viewpoint changes from the position of the less-exposure view to that of the more-exposure view is in the direction toward the occluded end of the stem at the less-exposure junction.*

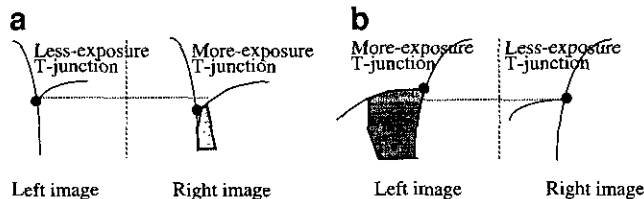


FIG. 6. Classification of T junctions in stereo views: (a) stem toward the right of cap; (b) stem toward the left of cap.

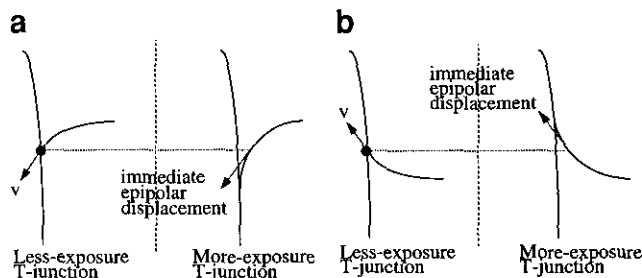


FIG. 7. Direction of epipolar displacement of T junctions.

Proof. A proof of the constraint in more formal terms can be found in Appendix A. ■

3.3.2. Limb Junctions

Lim and Binford [19] have made use of limbs but have presented no method to identify them. Identifying limbs is in fact a first step to the identification and description of curved surfaces. The appearance of each limb edge itself is no different from that of a real edge, but rather it is the junctions formed at the two ends of the limb edge which make the difference. Nalwa [26] has shown that the branches of a limb junction will be all cotangent at the junction. But identifying limb junctions in real pictures merely from this cue may be difficult.

A limb junction is formed when the line of sight from the camera is tangential to the contour at the intersection of the curved surface and its terminator surface (Fig. 8).

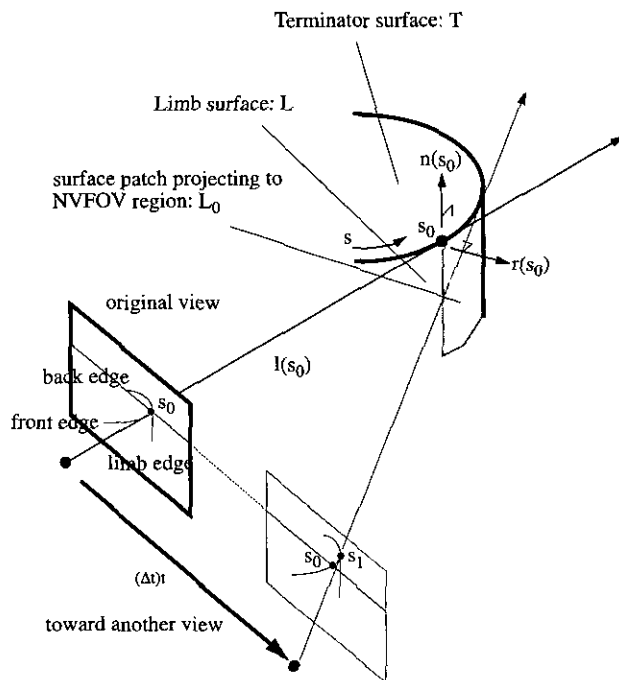


FIG. 8. Behavior of a limb junction upon change of the viewpoint.

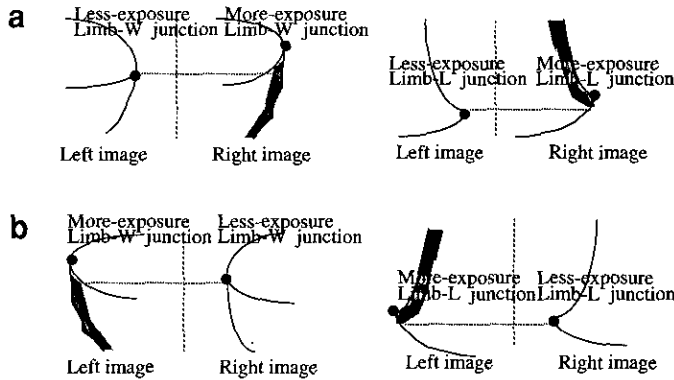


FIG. 9. Classification of limb junctions in stereo views: (a) back edge physically toward the right of limb edge; (b) back edge physically toward the left of limb edge.

It can be viewed as consisting of three branches: a limb edge, and two edges projected by the terminator contour. We will call the terminator edge closer to the camera the front edge, and the other terminator edge the back edge. Note that in a limb-W junction all branches are visible, and in a limb-L junction the back edge is occluded. The branches can be easily identified by checking their approximate disparities. The edges of the terminator surface are the branches closest to and farthest away from the camera, and the limb edge is in the middle.

From the perspective of occlusion, the front edge and the limb edge are from the occluding portion of the curved surface, and the back edge is from the occluded portion. We can thus tell whether a limb junction in a specific view is a less-exposure junction or a more-exposure junction by checking whether the back edge is physically toward the left or toward the right of the limb edge (Fig. 9). Note that due to the nature of the limb junction, the back edge is physically toward the right of the limb edge if it is on the left of the limb edge in the image. The back edge is invisible in a limb-L junction, but its position relative to the limb edge is revealed from that of the front edge.

As the viewpoint moves from the position of the less-

exposure view to that of the more-exposure view, the NVFOV region gradually appears and moves the limb junction along the projection of the terminator contour toward the back edge. As a result, the limb junction will go up or go down across the epipolar lines according to the orientation of the vector \mathbf{v} , where \mathbf{v} is the tangential vector to the projection of the terminator contour at the less-exposure junction and in the direction toward the back edge.

We thus have the following constraint:

THEOREM 2. LIMB-JUNCTION EPIPOLAR DISPLACEMENT CONSTRAINT. *A limb junction generally moves across the epipolar lines as the viewpoint changes, and the immediate epipolar displacement as the viewpoint changes from the position of the less-exposure view to that of the more-exposure view is in the direction toward the back edge at the less-exposure junction.*

Proof. A proof of the constraint in more formal terms can be found in Appendix B. ■

A limb junction is therefore one that has the shape of a W or L junction monocularly, and exhibits epipolar displacement across stereo views according to the above constraint (Fig. 10). This presents an additional cue to identify the limbs. The requirement is that at least one end of the limb must terminate at a terminator surface. But we conjecture that without presence of other cues such as shading and texture, limbs not terminating at terminator surfaces are not identifiable from binocular stereo alone; e.g., the stereo images of a sphere are not distinguishable from those of a circular plate.

3.3.3. Edge Junctions

NVFOV surfaces (Fig. 11) along orientation discontinuities are usually ignored in stereo processing systems. As surfaces composing an edge junction must be from the same object, each of these NVFOV surfaces shares the same object with the rest of the surfaces around the edge junction. They are therefore important for object level

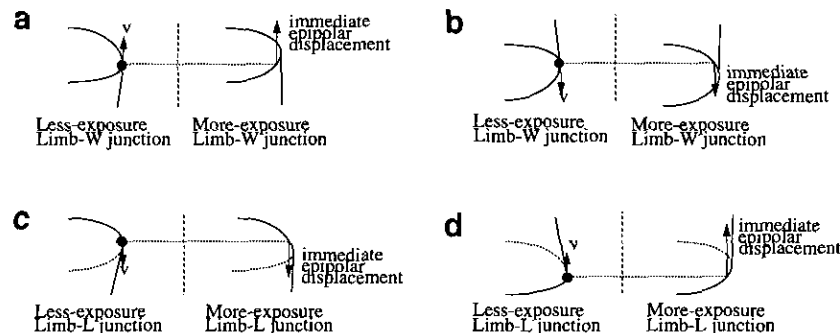


FIG. 10. Direction of epipolar displacement of limb junctions.

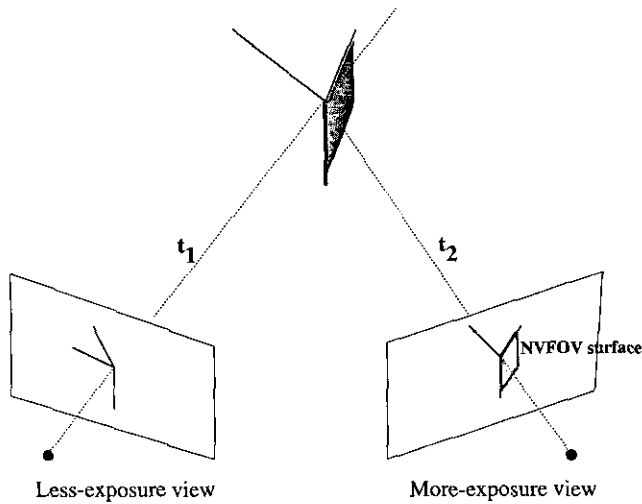


FIG. 11. Formation of NVFOV surface at orientation discontinuity occlusion.

description and recognition. Although a few attempts [13, 17] have been made to infer about these NVFOV surfaces, they apply to specific domains such as architectural scenes only. Domain-specific knowledge, such as orthogonal trihedral vertices [17] and restricted vanishing points of all lines in the scene [13], are employed. We seek to extend the concept to cover generic objects.

To detect these NVFOV surfaces, we can check for any change in the junction type of the corresponding edge junctions across stereo views. As surfaces are occluded in one of the views, the corresponding junctions will also appear to be surrounded by different number of surfaces.

Moreover, as surfaces composing an orientation discontinuity which is concave to the viewing direction of an image cannot occlude each other in that image (if one is visible, then both are visible), orientation discontinuity occlusions can only happen around convex edges. This also means orientation discontinuity occlusions can only happen around convex vertices, as the convexity of a trihedral vertex is uniquely defined by the convexity of any one of the composing edges. We therefore have an additional constraint to detect orientation discontinuity occlusions:

CONVEXITY CONSTRAINT. *The correspondence of the edge junctions (projected from a trihedral vertex) around an orientation discontinuity occlusion must recover a convex vertex, not a concave one.*

The change in the junction type is thus asymmetric across stereo views. While a W junction in the left view is matchable with a Y junction in the right view, as their matching returns a convex vertex, a Y junction in the left view is not matchable with a W junction in the right view. Examples of permissible change in the junction type are given in Fig. 12.

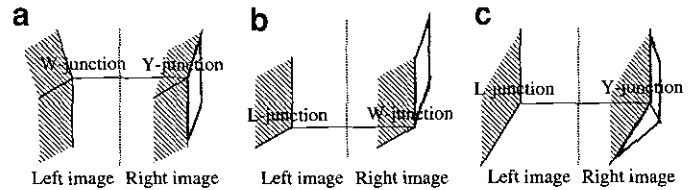


FIG. 12. Examples of permissible change of the junction type at orientation discontinuity occlusions for trihedral objects.

4. A HIERARCHICAL FEATURES-BASED APPROACH

Most stereo systems match very local features (such as intensities or edgels) or match extended features such as line segments. To resolve global ambiguity, they employ constraints of surface continuity of disparity and preservation of the order of features along epipolar lines. These constraints, however, are valid only in the interior of a smooth surface and not across the occlusion boundaries. The traditional approach is to extract the occlusion boundaries from the stereo disparity data, but we argue that this approach already incorporates errors that cannot be corrected later.

Instead, we propose to pursue an approach that computes descriptions at a hierarchy of levels with matching also done at various levels. This approach has many advantages. Higher level features are more distinct and easier to match (the complexity is lower and the errors are fewer). Furthermore, identifying surfaces and occlusion boundaries allows us to apply the smoothing and ordering constraints only to those regions where they are likely to hold. The figural continuity constraint is built in. Lower level features, matched in the context of higher level matches, can help provide higher accuracy and also allow us to use texture and other surface markings that do not form any higher level structures.

In our system we use the following hierarchy of features: edges, curves, ribbons, and junctions. Note that junctions are not only highly localized and distinct features, but matching them also gives us a means of identifying the different occlusions described in Section 3.3.

The first problem in using the described feature hierarchy is the difficulty of computing them from monocular images of real scenes. Boundaries found by edge detection techniques tend to be fragmented and several edges that do not correspond to object boundaries, but are caused by surface markings, highlights, and noise, are likely to be present. To extract surfaces from such features is a difficult task. We use a perceptual grouping technique described by Mohan and Nevatia [24, 25]. In this approach relationships of cocurvilinearity and symmetry are employed to extract collated features of various abstractions from the image. This method, along with some of our modifications, is described briefly in Section 4.2.

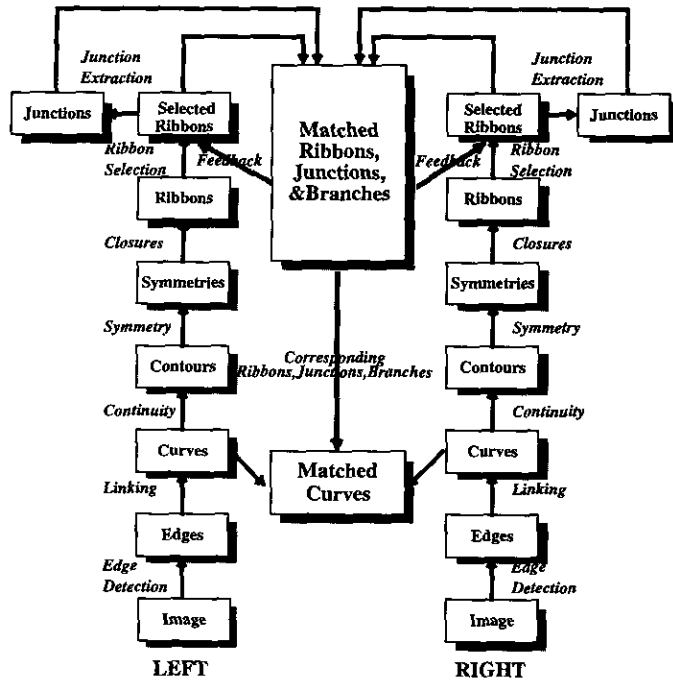


FIG. 13. Overview of our stereo system.

The monocular grouping algorithm is highly effective but not expected to be perfect. We allow feedback between the stereo matching process and the ribbon selection process, so that we can get better groupings by the use of two views that we could by each taken separately.

A block diagram of our system is shown in Fig. 13. Note that there are effectively three subsystems:

1. Monocular groupings of the left and right views;
2. Stereo correspondence; and
3. Feedback from stereo correspondence to monocular groupings.

Many steps of the system require satisfying a given set of constraints. In our system, constraint satisfaction is done by a relaxation network as described below. After this, we describe each subsystem one by one.

4.1. Constraint Satisfaction Methodology

In many parts of our system, features are related by a set of constraints. We model all constraints as being either *unary* or *binary*. Unary constraints come from individual merits of a feature. Binary constraints relate a pair of features. We further subclassify unary and binary constraints according to whether the constraint is absolute or tolerable:

1. Unary Constraint

(a) *Unary Absolute Constraint*. It is a unary constraint that must be satisfied.

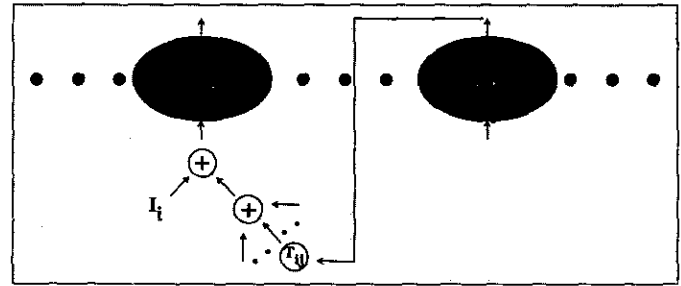


FIG. 14. A simple relaxation network satisfying unary and binary constraints.

(b) *Unary Excitatory or Inhibitory Constraint*. It is a unary constraint that represents how good (excitatory) or how bad (inhibitory) a feature is in a certain aspect. It will be positive if it is excitatory, negative if it is inhibitory.

2. Binary Constraint

(a) *Binary Mutually Exclusive Constraint*. Two features are mutually exclusive if at most one of them can survive at the same time.

(b) *Binary Excitatory or Inhibitory Constraint*. It is a binary constraint that represents whether two features support (excitatory) or conflict with (inhibitory) each other according to a certain aspect. It will be positive if it is excitatory, negative if it is inhibitory.

In the rest of this chapter, as we mention various constraints for monocular grouping or stereo correspondence, we will indicate their type according to the above terminology.

This formulation of constraints as either unary or binary naturally defines a network with the features or the potential matches between features serving as the nodes and their relationships serving as the arcs. To satisfy the unary and binary constraints, we define a modification of the Hopfield network [15, 16]. The modification is guaranteed to ensure that absolute constraints are satisfied exactly, which is not necessarily the case with Hopfield network. This relaxation network is applied throughout our stereo system in the selection and stereo matching of features at various levels and is described below.

4.1.1. A Simple Relaxation Network

Figure 14 shows the relaxation network for selection among a set of nodes from node 1 to node N . Each node is a feature or a potential match between two features depending upon the application, and it must satisfy all the unary absolute constraints. The aim of the network is to select the nodes which are the best among all their competitors, subject to merits both from themselves and through support from the others.

Take any node i among all the nodes. It has an input

U_i and an output V_i . The output V_i can be 0 or 1; 1 means selected, and 0 means rejected. Input U_i is the sum of two entities, I_i and $\sum_{j=1}^N T_{ij}V_j$, which represent respectively how good node i itself is and how much total support node i obtains from the other nodes in the network. To be precise, I_i is the weighted sum of the unary excitatory and inhibitory constraints for node i , whereas T_{ij} is the weighted sum of the binary excitatory and inhibitory constraints between node i and any other node j in the network,

$$I_i = \sum_u w_u C_u(i),$$

where $C_u(i)$ is the u th unary excitatory or inhibitory constraint for node i , and w_u is the corresponding weight;

$$T_{ij} = \sum_b w_b C_b(i, j),$$

where $C_b(i, j)$ is the b th binary excitatory or inhibitory constraint between node i and any other node j , and w_b is the corresponding weight.

The transfer function from U_i to V_i is a *winner-takes-all* function; i.e., in every iteration, among all the nodes which are mutually exclusive with node i , only the node which has maximum U value will have its output V set to 1. The rest will have their outputs reset to 0. To sum up, the dynamics of the network is governed by two equations:

$$U_i = \sum_{j=1}^N T_{ij}V_j + I_i,$$

$$V_i = \begin{cases} 1 & \text{if } U_i = \max(U_s) \forall s \text{ s.t.} \\ & \text{node } s \text{ is mutually exclusive with node } i \\ 0 & \text{otherwise.} \end{cases}$$

It is therefore guaranteed that any two nodes being mutually exclusive with each other will never come out from the relaxation at the same time. The network is also capable of delivering maximally allowable number of output nodes which are not contradictory to each other.

The network is iterated until the outputs of the nodes reach an equilibrium state or the number of iterations exceeds an upper limit. The outputs of the relaxation are the nodes which have their output values V being 1.

Note that if the binary mutually exclusive constraint is *transitive*, i.e., if nodes i and j are competitors and nodes j and k are competitors then nodes i and k must be competitors, then the nodes can be split into distinct groups of competing nodes and each iteration can be done in parallel among the groups. If there is no mutually exclusive constraint, each iteration will be totally parallelizable among all the nodes.

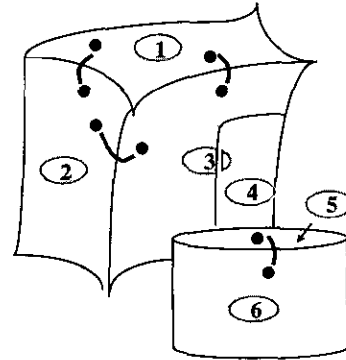


FIG. 15. Some constraints for ribbon selection. (a) Uniqueness constraint: ribbons 3 and 4 are in absolute conflict; (b) solid-formation constraint: ribbons 1-3, and ribbons 5 and 6 form two 3-D objects respectively.

Note also that our approach is not limited to the above simple relaxation network. It is our belief that exhaustive search among low-level features is computationally infeasible, but as we go to higher level features, the features are more distinct and fewer in number, thus any simple mechanism will suffice to do the selection or matching job.

4.2. Monocular Perceptual Grouping

Mohan and Nevatia [25] have proposed using relationships of cocurvilinearity and symmetry to extract collated features of various abstractions from each image (Fig. 13). *Edges* are detected from each image using Canny's edge-detector [4] and are linked into edge contours using eight-neighbor connectivity. Edge contours are segmented into *curves* [28] at *curvature extrema* so that every curve is smooth in itself, and curves are grouped into *contours* based on cocurvilinearity. *Symmetries* are then detected from each pair of approximately symmetrical contours, and they form *ribbons* if they have proper end closures at both ends of the symmetries. Closure at the end of a symmetry can be composed of a contour, a set of multiple contours, or the ends of other symmetries. A number of conflicting ribbons are computed and selection of ribbons must be done to resolve conflicts.

We believe ribbons should not be selected merely based on their individual merits. If objects in the scene are all composed of thin plates, that is the only solution as the surfaces are indeed truly independent. But surfaces within a 3-D object do exhibit certain relationships with one another. Note that most of the objects in real life are three-dimensional, e.g., a desk, a cup, and a TV set, and they are the objects that humans are most interested in recovering their 3-D descriptions.

The basic relationship between two adjacent smooth surface patches within a 3-D object is that they share a smooth contour which is bounded by non-T junctions only (Fig. 15b). We implement this relationship as a strong excitatory constraint in the ribbon-selection process.

TABLE 1
Constraints for Ribbon Selection

Constraint	Nature	Remarks
Symmetry smoothness (SSC)	Unary, +	A ribbon with smoother symmetry axes is more likely to be a real ribbon. Smoothness is measured in terms of number of corners in our implementation.
Boundary evidence (BEC)	Unary, +	A ribbon with less fragmented boundary is more likely to be a real ribbon.
Uniqueness	Binary, ME	As opaque surfaces are assumed, each point visible in an image can belong to at most one surface. Two ribbons having overlap in their contained points are therefore in absolute conflict (Fig. 15a). However, we do allow two ribbons with one totally embedded inside the other to exist together, as we regard them as one on top of the other.
Solid-formation (SFC)	Binary, +	Two ribbons which are not in conflict and share a smooth contour bounded by non-T junctions support each other, as they are likely to be projected from adjacent surfaces of the same 3-D object (Figure 15b).

Note. +, excitatory; -, inhibitory; A, absolute; ME, mutual exclusive.

We thus have the constraints outlined in Table 1 for selection of ribbons.

As there are altogether three compromisable constraints, two predefined weights in the weighted sum of the constraints are necessary for the cost function to be optimized in the relaxation. The cost function is therefore

$$E(V) = -\sum_i V_i(\text{BEC}_i + w_{\text{SSC}}\text{SSC}_i) - \frac{1}{2} \sum_{ij} V_i V_j w_{\text{SFC}}\text{SFC}_{ij},$$

where w_{SSC} and w_{SFC} are kept constant at 1 and 3 respectively in our system. The values are designed according to the relative importance of the constraints.

Once surfaces have been segmented and decomposed into component ribbons, extraction of junctions from each image becomes an easy task. Owing to the assumption that vertices are at most trihedral, we only extract L, W, Y, and T junctions. L and W junctions might turn out to be limb-L and limb-W junctions in stereo matching.

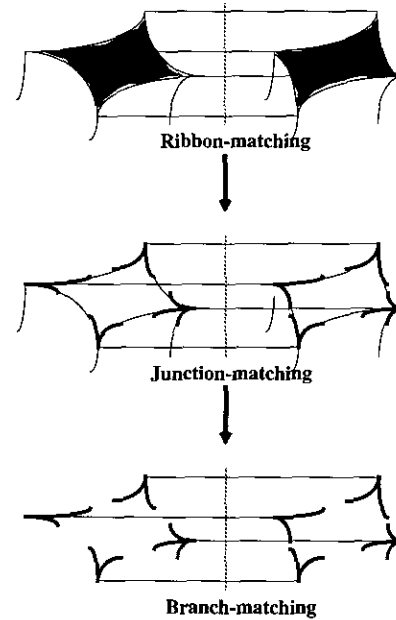


FIG. 16. Hierarchical matching.

4.3. Binocular Stereo Correspondence

Matching of features follows a hierarchical order. Two ribbons are potentially matchable only if all their junctions are matchable. Two junctions are potentially matchable only if all their branches that belong to the ribbons being matched are matchable (Fig. 16). We will present the physical constraints for matching the features at various levels and conclude about how different types of occlusions are identified.

4.3.1. Constraints for Ribbon Matching

Ribbons in the left and right views are matched using the constraints outlined in Table 2 and shown in Fig. 17.

As there is only one compromisable constraint, no weight in the weighted sum of the constraints is necessary for the cost function to be optimized in the relaxation.

4.3.2. Constraints for Junction Matching

In our system, junction matching is initiated by ribbon matching. To consider possible matching between two junctions, only the branches which belong to the two ribbons being matched are involved. Junctions are matched using the constraints outlined in Table 3 and shown in Figs. 18 and 19.

As there is only one compromisable constraint, no weight in the weighted sum of the constraints is necessary for the cost function to be optimized in the relaxation.

In matching viewpoint-dependent junctions (T junctions and limb junctions) across stereo views, we must make sure the direction of "immediate" epipolar displacement

TABLE 2
Constraints for Ribbon Matching

Constraint	Nature	Remarks
Epipolar	Unary, A	Two ribbons are potentially matchable only if they have corresponding bounds of vertical extent.
Shape correspondence	Unary, A	Two ribbons are potentially matchable if all their junctions other than those between the T junctions are matchable (Fig. 17a). To take into account possible discrepancies in detecting junctions from the two views, highly obtuse L junctions can be exceptions.
Uniqueness	Binary, ME	Two ribbon matches are in conflict if there is any conflict between their proposed junction matches.
Figural continuity	Binary, +	Two ribbon matches support each other if they are not in conflict and there is overlap in their proposed junction matches (Fig. 17b).

Note. +, excitatory; -, inhibitory; A, absolute; ME, mutual exclusive.

satisfies the appropriate junction epipolar displacement constraint. But how do we check the direction of immediate epipolar displacement in the first place?

The immediate displacement v and the overall displacement V may not have the same direction (but this happens only under extreme cases when the views are widely displaced and there is a change of convexity of the involved contour at the junction point during the viewpoint change), as shown in Fig. 20.

To examine if two junctions in stereo views satisfy the immediate displacement direction constraint, we can first check if the junction point J in the less-exposure view has a point match P' in the more-exposure view along the

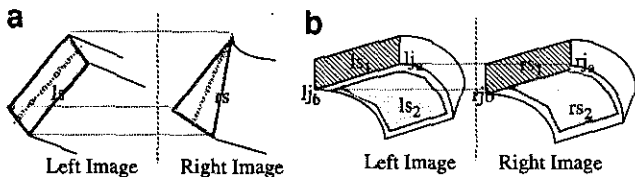


FIG. 17. Some constraints for ribbon matching: (a) Shape-correspondence constraint: ribbons ls and rs have similar axes of symmetry and similar vertical extents, yet they are not matchable as not all their junctions are matchable. (b) Figural-continuity constraint: ribbon-matches (ls_1, rs_1) and (ls_2, rs_2) support each other as they propose same junction matches (lj_a, rj_a) and (lj_b, rj_b).

TABLE 3
Constraints for Junction Matching

Constraint	Nature	Remarks
Type	Unary, A	A non-T junction can be matched with a non-T junction, provided that they satisfy the convexity constraint described in Section 3.3. A more-exposure T junction can only be matched with a less-exposure T junction. To take into account of possible disocclusion of T junction (Fig. 18), we also allow a less-exposure T junction to be matched with a non-T junction, but not with another less-exposure T junction.
Epipolar	Unary, A	Two junctions are potentially matchable only if they lie on corresponding epipolar lines. This constraint is relaxed for T junctions and possible limb junctions, as their counterparts generally do not fall on corresponding epipolar lines. However, the direction of immediate epipolar displacement must follow the T junction and the limb junction epipolar displacement constraints described in Section 3.3.
Shape correspondence	Unary, A	Two junctions are potentially matchable only if all their branches that belong to the ribbons being matched are matchable.
Uniqueness	Binary, ME	A non-T junction can be matched with at most one non-T junction, but it can be matched with more than one less-exposure T junctions if it is occluded in the other view (Fig. 18).
Figural continuity	Binary, +	If junction l is matched with junction r , it is preferred that junctions which share branches with junction l and junction r in the two views are also matched, if those junctions are indeed matchable (Fig. 19).

Note. +, excitatory; -, inhibitory; A, absolute, ME, mutual exclusive.

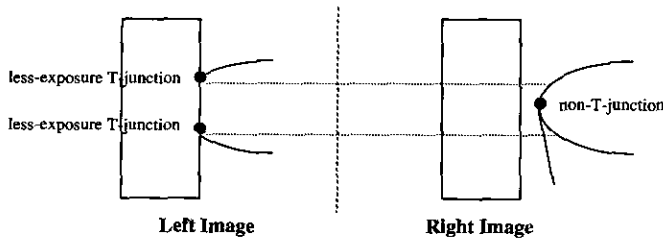


FIG. 18. Type and uniqueness constraints for matching junctions. (a) Type constraint: a less-exposure T junction can be matched with a non-T junction. (b) Uniqueness constraint: more than one less-exposure T junction can be matched with a non-T junction.

corresponding contour C' (how to determine whether a certain junction in a certain view is less-exposure or more-exposure is described in Section 3.3; it can be determined from the relative positions of the junction branches). If the two junctions are indeed matchable, such a point P' must exist, as we are referring to a viewpoint change from the less-exposure view to a more-exposure one. Then the immediate displacement direction \mathbf{v} is the tangential vector to the contour C' at the point P' in the direction leading along the contour to the more-exposure junction J' .

The method applies to both T junctions and limb junctions. Examples of T-junction and limb-junction correspondences violating the immediate displacement direction constraint are shown in Fig. 21.

However, extracting the direction of the immediate displacement requires point-to-point correspondences between the involved contours. This is not as trivial as it seems to be, as a contour point in one view may have more than one possible match in the other view. For example, the point at junction J shown in Fig. 20 can be matched with either point P' or point Q' in the more-exposure view. We use a method described in Section 4.6.

4.3.3. Constraints for Branch Matching

Branches around matchable junctions are matched using the constraints outlined in Table 4 and shown in Fig. 22.

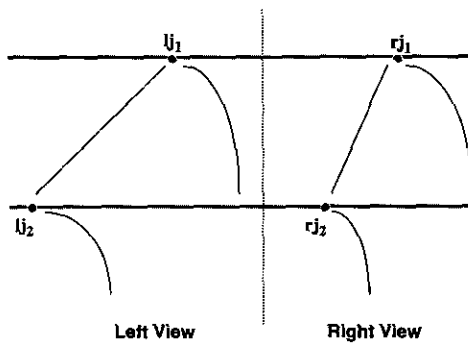
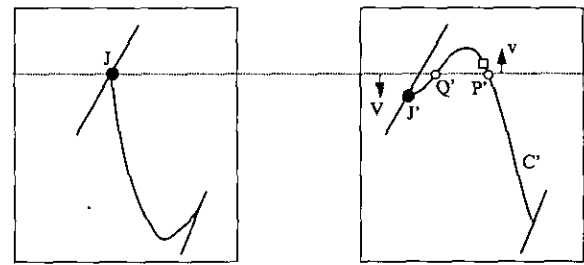
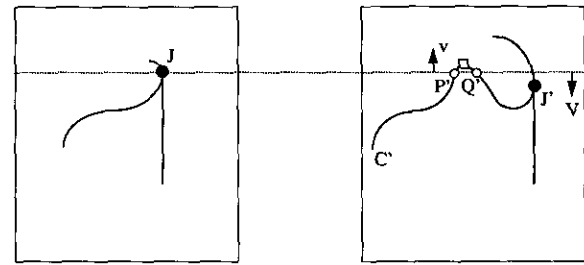


FIG. 19. Figural continuity constraint for matching junctions: matches (lj_1, rj_1) , (lj_2, rj_2) support each other.



T-junction correspondence



Limb-junction correspondence

FIG. 20. The immediate displacement direction (\mathbf{v}) and the overall displacement direction (\mathbf{V}) of a viewpoint-dependent junction across epipolar lines in stereo views.

As there is no compromisable constraint, no weight in the weighted sum of the constraints is necessary for the cost function to be optimized in the relaxation.

4.3.4. Identification of Occlusions

Matching surface level features automatically removes the intersurface disparity discontinuity problem in stereo correspondence, but the intrasurface stereo noncorrespondence problem still must be resolved by identifying the occlusions.

Different types of occlusions can be identified before or during stereo correspondence using concepts presented in Section 3.3. Their identifications are based on the identifications of junctions:

1. T junctions are identified monocularly from the selected ribbons in each view. False boundaries of the occluded surfaces, which are portions of their apparent boundaries terminating at T junctions, will be excluded from stereo matching.
2. Limb junctions are W or L junctions whose branches are approximately cotangent and which present epipolar displacements across stereo views. Limb edges, which are edges terminating at limb junctions, will be excluded from stereo matching as well.
3. Edge junctions changing their junction type across

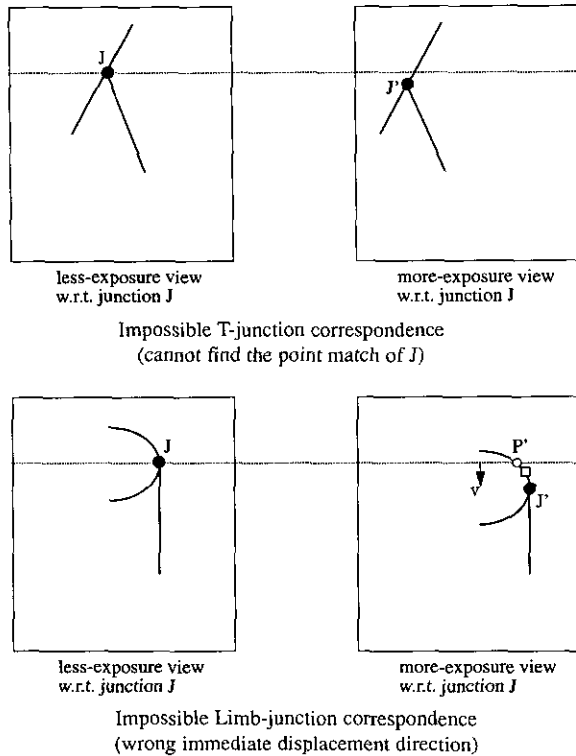


FIG. 21. Junction matches violating the immediate displacement direction constraint.

stereo views and recovering convex vertices will mean occurrence of NVFOV surfaces from orientation discontinuity occlusions. The NVFOV surfaces are surfaces around the edge junctions that remain unmatched in ribbon-matching.

4.4. Feedback from Stereo Correspondence to Monocular Groupings

The previous section described how stereo matching is carried out for the hierarchical features selected from the left and right views. As interpretations from the separate views may not totally agree, there are ribbons which are selected in one view but their corresponding ribbons are not selected in the other view (Fig. 23). Such ribbons would not be matched in the stage of stereo correspondence.

The ribbons which are matched are first put into a group which we call the *consistent ribbon pairs*, as they are the ones where current interpretations of separate views are consistent. Note that NVFOV surfaces along orientation discontinuity occlusions will not and should not be matched. They are put into the same group of the *matched ribbons*.

For the rest of the selected ribbons in each view, we look at the pool of all possible ribbons in the other view and search for potentially matchable ribbons which are

TABLE 4
Constraints for Branch Matching

Constraint	Nature	Remarks
Epipolar	Unary, A	Two branches are potentially matchable only if the tangential vectors along them are either both point up or both pointing down across the epipolar lines (Fig. 22a). If the tangential vectors are approximately parallel to the epipolar lines, they must be both pointing toward the left or toward the right.
Uniqueness	Binary, ME	One branch can be matched with at most one branch in the other view.
Surface orientation	Binary, ME	The branch matches associated with a junction match must be such that the surface normals of the corresponding surfaces have their z components n_z 's (components perpendicular to the reference image plane) in the same direction. This implies that branches all pointing up or all pointing down across the epipolar lines must be matched in the order from left to right (Fig. 22b).

Note. +, excitatory; -, inhibitory; A, absolute; ME, mutual exclusive.

not in conflict with any of the consistent ribbon-pairs. If there is no such a ribbon from the other view, the unmatched ribbon will be rejected. If there is, the two ribbons will be a potential match. We will call all these potential matches *inconsistent ribbon pairs*.

Inconsistent ribbon pairs undergo the process of ribbon selection again. This time each node in the relaxation network is not a single ribbon, but instead a ribbon pair from the left and right views respectively. The relaxation process is similar to that in the first ribbon selection, except that merits of the two ribbons in a node due to various constraints are averaged to become the merits of the node. Output values of the nodes representing the consistent ribbon pairs are always 1; i.e., they are always selected, but they will be involved in giving binary constraints to the nodes representing the inconsistent ribbon pairs. Outputs of the relaxation are then ultimate pairs of selected ribbons from the left and right views of the scene.

As the relaxation network always gives the maximally allowable number of nodes which do not have conflict, two cycles of ribbon selection can therefore average the groupings from the left and right views and give a consistent interpretation of the scene.

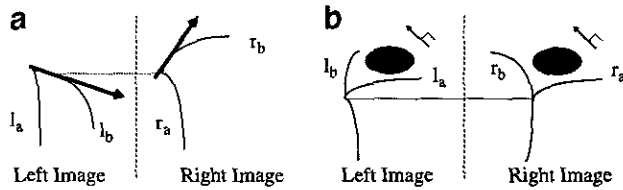


FIG. 22. Some constraints for branch matching: (a) Epipolar constraint: the two junctions are not matchable as none of l_a and l_b is matchable with r_b ; (b) surface-orientation constraint: l_a must be matched with r_a and l_b must be matched with r_b , so that $l_a \times l_b$ and $r_a \times r_b$ point to the same direction (out of paper).

4.5. Stereo Correspondence of Surface Markings

As ribbons, junctions, and branches of junctions are hierarchically matched, the boundary contours of each ribbon, which terminate at junctions, are also matched automatically. In particular, depth measures along boundary edges *parallel to the epipolar lines*, which are ignored in most stereo work, can also be interpolated from the depth measures of the junctions at which they terminate.

In our approach, surfaces in the scene are segmented in the process of stereo computation. This may be sufficient for many description and recognition purposes. A denser disparity map can also be obtained by matching the surface markings on the surfaces. We choose to match curves rather than edgels because of their higher abstraction. Two curves are potentially matchable only if they are within corresponding ribbons. As disparities along the boundary of each ribbon have been obtained, on assuming ribbons

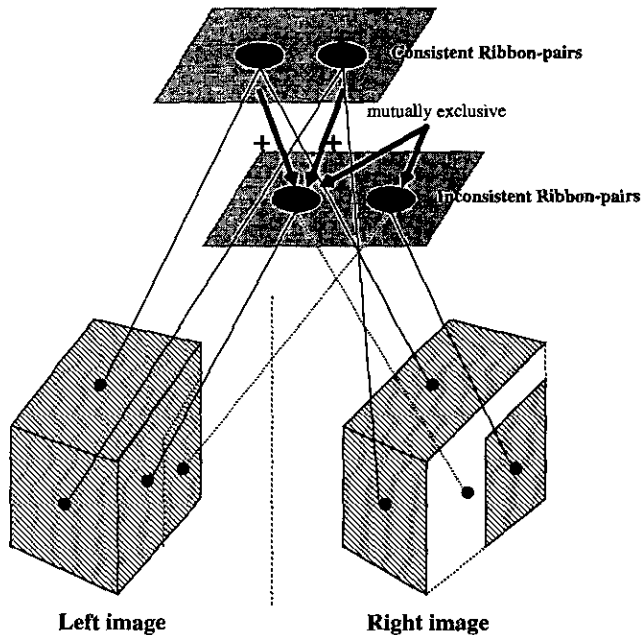


FIG. 23. Consistent and inconsistent ribbon pairs.

TABLE 5
Constraints for Curve Matching

Constraint	Nature	Remarks
Epipolar	Unary, A	Curves are potentially matchable only if they have overlap in their vertical extents across the epipolar lines.
Vergence (VC)	Unary, +	A match between two curves is more preferred if their disparity is closer to the estimated disparity on the corresponding ribbons.
Uniqueness	Binary, ME	A curve cannot be simultaneously matched with two curves in the other view which have overlap in their vertical extents.
Ordering	Binary, ME	Curve matches must follow the order from left to right.
Surface continuity (SCC)	Binary, +	Two curve matches with their curves next to each other in the same order support each other.
Figural continuity (FCC)	Binary, +	Two curve matches with their curves being connected to each other support each other.

Note. +, excitatory; -, inhibitory; A, absolute; ME, mutual exclusive.

are planar, we can estimate the disparity of a curve at any epipolar line from its position inside its own ribbon. We can also employ all the constraints generally used in stereo correspondence which are applicable for features within a surface. The constraints for curve-matching are given in Table 5.

As there are altogether three compromisable constraints, two predefined weights in the weighted sum of the constraints are necessary for the cost function to be optimized in the relaxation. The cost function is therefore

$$E(V) = -\sum_i (V_i V C_i) - \frac{1}{2} \sum_{ij} V_i V_j (w_{SCC} SCC_{ij} + w_{FCC} FCC_{ij}),$$

where w_{SCC} and w_{FCC} are kept constant at 0.5 and 1 respectively in our system. The values are designed according to the relative importance of the constraints.

4.6. Recovering Disparities of the Matched Features

Once the ribbon boundaries and the texture markings are matched, we will have to set up point-to-point correspondences between the counterparts to recover the disparity measure. This is not simple, as a contour point in

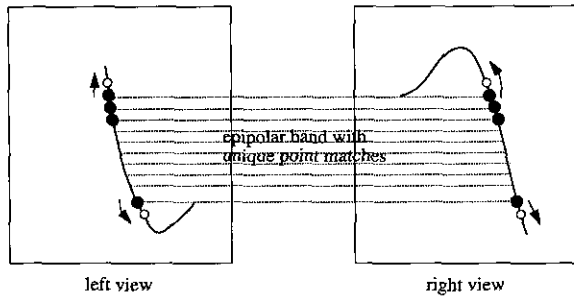


FIG. 24. Determining point correspondences between two corresponding contours in stereo views.

one view may have more than one possible match in the other view along the corresponding epipolar line.

Here we use a simple method to set up point correspondences between two corresponding contours. It is based on an assumption: there exists at least one unique point correspondence between the two corresponding contours based on the epipolar constraint alone; i.e., there exists at least one pair of corresponding epipolar lines in the two views such that only one point from each contour falls on the lines. As shown in Fig. 24, we first locate such unique point matches, and from them we grow the rest of the point correspondences by extending the correspondences both up and down across the epipolar lines, based on the continuity constraint. However, there may still be ambiguity in growing the point matches, especially for those closed contours as shown in Fig. 25. Since the ordering constraint strictly holds for all the points on the same surface, we just match the points according to the left-to-right order. For example, points P and Q are matched with points P' and point Q' respectively in Fig. 25.

5. EXPERIMENTAL RESULTS

We show results on different kinds of scenes to illustrate the capability of our approach. The simple relaxation network described in Section 4.1.1 is applied throughout our stereo system in the selection and stereo matching of fea-

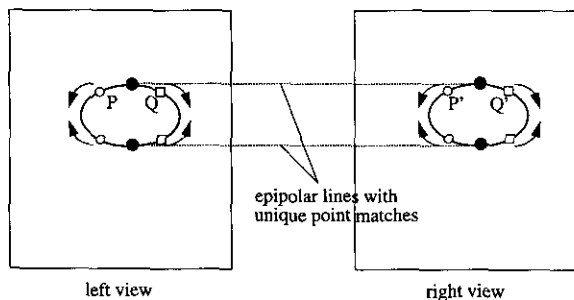


FIG. 25. Determining point correspondences between two closed contours in stereo views.

tures at various levels. All relaxations are found to converge within 20 iterations in all test images.

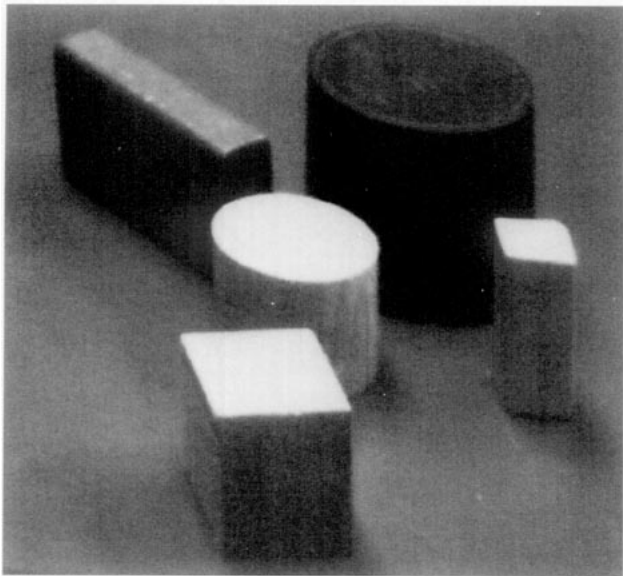
Figure 26 shows the images, the hierarchical features, and the stereo matching results of a scene with multiple occlusions. All the ribbons, junctions, and contours are correctly matched. In Fig. 27, the correspondences of the T junctions in the left view are also shown, just to present some of the junction matches. Note that multiple less-exposure T junctions may be matched together with a non-T junction, as there may be disocclusion upon change of viewpoint, and this is stated in the Type and the Uniqueness Constraints outlined in Table 3. Because of this and the large number of junctions involved, the task of junction matching is difficult. We have experienced matching errors when junctions are hypothesized separately from each view and matched based on the junction-matching constraints alone, as local information may not be enough to resolve all the correspondence ambiguity. However, the hierarchical nature of the system allows the matching to be guided by ribbon matching at a higher level which contains more global information, and thus ambiguity is resolved.

Note also that in this example there are many boundary edges which are parallel to the epipolar lines, and this is a problem for most stereo systems which are local feature based. However, since our system has the junction level in the feature hierarchy used, precise depth measures of the junctions at the ends of those boundary edges can be recovered. The system can then interpolate about the disparities of the boundary edges linearly from those of the junctions.

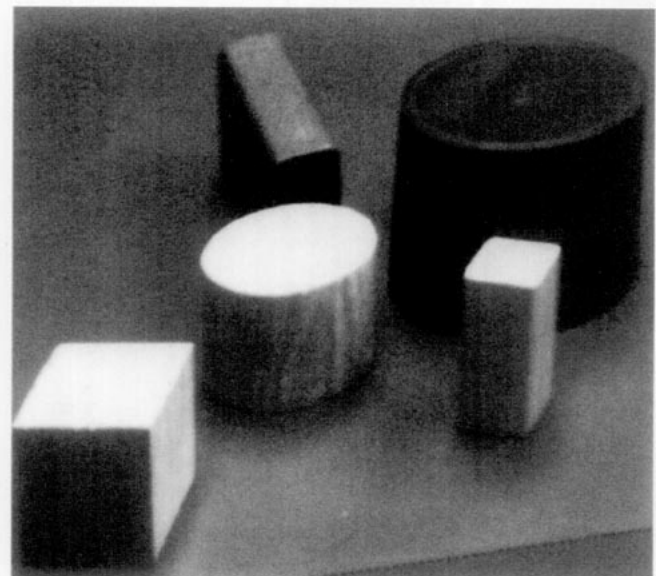
If the surface boundaries are matched blindly without identifying their natures, the depth map will be that as shown in Fig. 28a. Limb edges in stereo views in fact do not correspond to each other. Our stereo system is capable of identifying the limb edges from the limb junctions at which they terminate, and computing their depths at those junctions. Here for display purposes we assume disparity along a limb edge is linear between the limb junctions, and interpolate for the disparity measures along all the limb edges; the revised depth map is shown in Fig. 28b.

In the same figure we also show the NVFOV surfaces identified along the orientation discontinuity occlusions, for display purpose their disparities being interpolated as if they are planar.

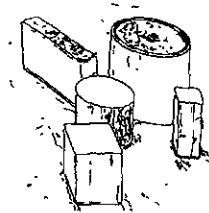
Another example is another scene with curved objects. Processing results of the scene are shown in Fig. 29. Note that selected ribbons from the monocular views are not consistent (Figs. 29g and 29h), and this is fixed by the feedback from stereo correspondence to monocular groupings (Figs. 29i and 29j). Figure 29k shows the complete disparity output, including the noncorrespondence of limb edges of the tape dispenser and the disparities of the surface markings. Here the limbs of the cylindrical object at the back are not identified because of lack of resolution.



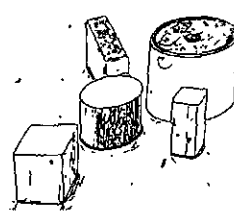
left image



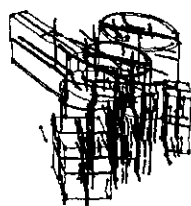
right image



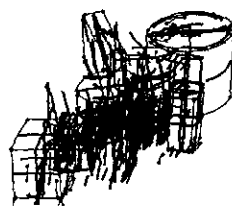
left edges



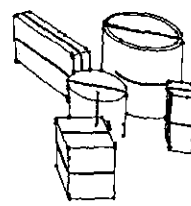
right edges



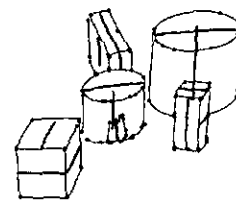
left symmetries



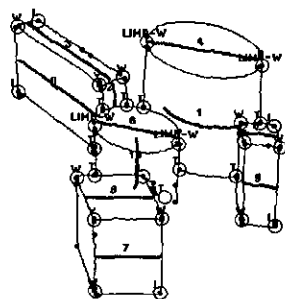
right symmetries



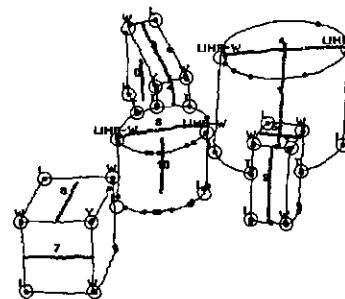
left selected ribbons



right selected ribbons



matched ribbons, junctions (left)



matched ribbons, junctions (right)

FIG. 26. Hierarchical features of a scene with multiple occlusions.

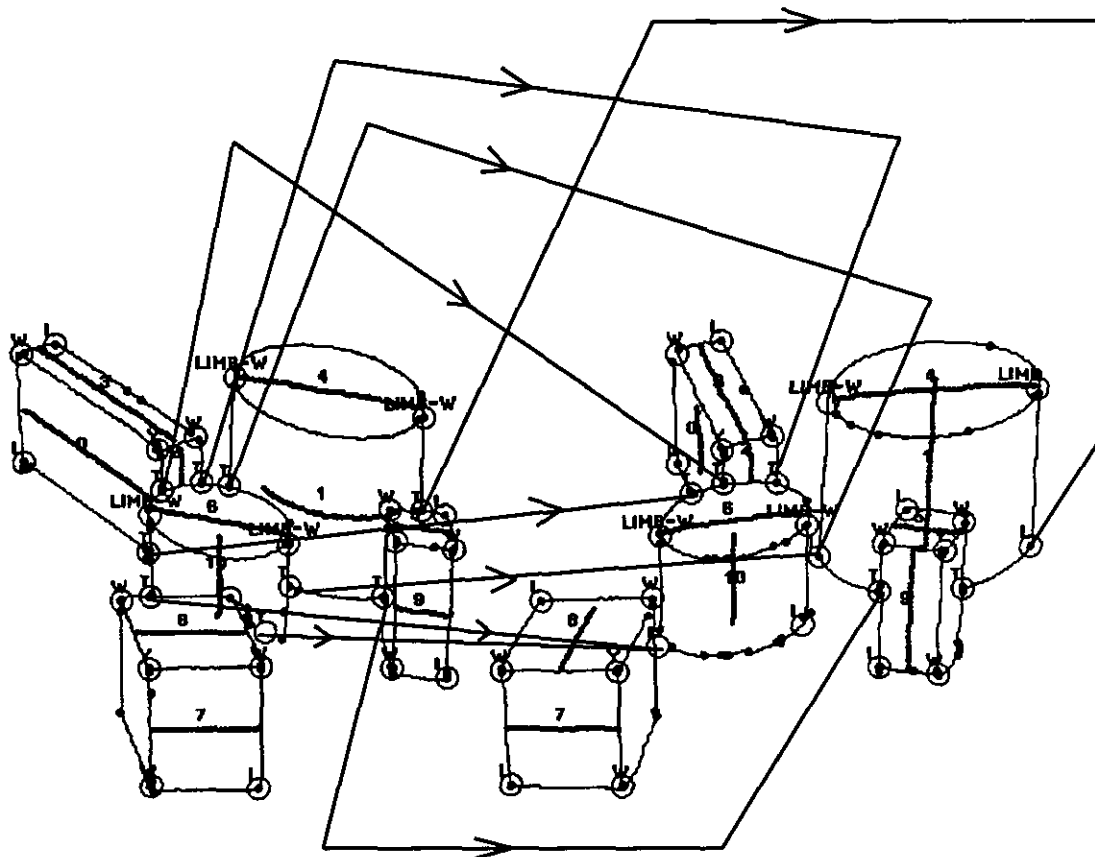


FIG. 27. Correspondences of T junctions in the left view of the scene with multiple occlusions.

Our stereo system requires higher resolution of the images for more accurate depth estimations.

The last example is a scene with significant texture. The images, the edges, and the disparity output are shown in Fig. 30. All the ribbons, junctions, and contours with the exception of a few in the background, are correctly matched. What is different in this example is that there are a lot of spurious false T junctions created by surface markings. Our system is able to distinguish the real T junctions (viewpoint-dependent) from the false ones (viewpoint-independent), as real T junctions move across epipo-

lar lines upon change of viewpoint (in a direction specified by the T Junction Epipolar Displacement Constraint described in Section 3.3.1) whereas false ones do not. On the junction-matching side, the set of highly restrictive junction-matching constraints (described in Table 3) plus guidance from the ribbon matching results at a higher level have been found to be very effective in ruling out wrong matches.

6. CONCLUSION AND FUTURE WORK

We have presented a stereo system that computes hierarchical descriptions of a scene from each view and combines the information from the two views to give a 3-D description of the scene. This system utilizes bilateral communication between monocular groupings and stereo correspondence. The hierarchy of descriptions helps reduce the computational complexity without sacrificing accuracy and helps avoid the errors caused by improper application of commonly used stereo correspondence constraints of surface continuity and ordering. Occlusions are specifically identified and interpreted. Visible surfaces are segmented and we can infer some properties of the surfaces visible in only one view. We are currently investigating how sur-

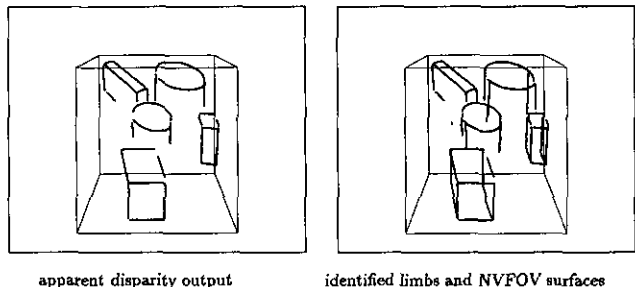
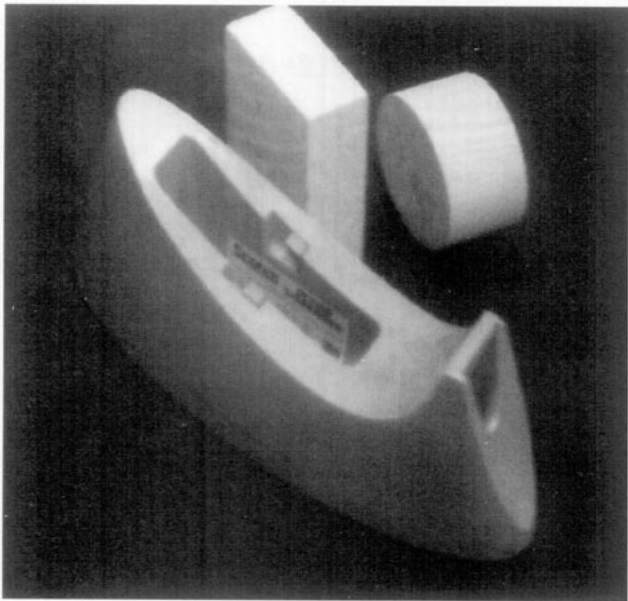
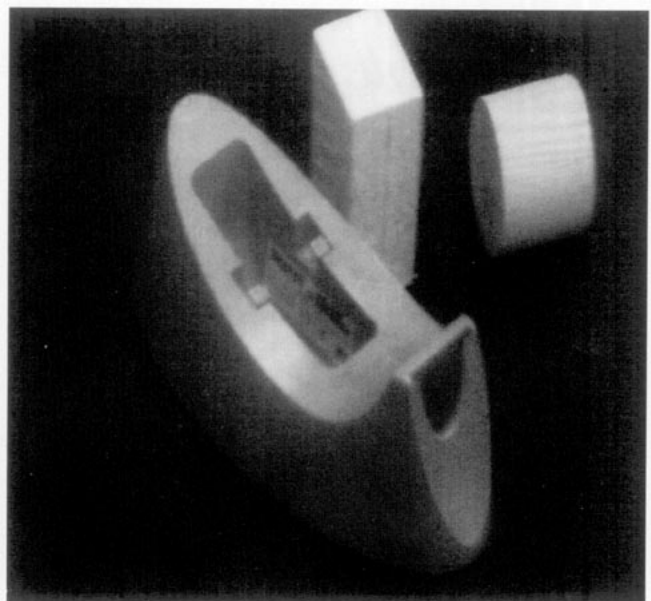


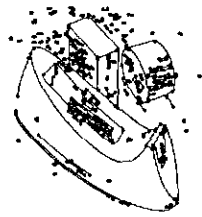
FIG. 28. Disparity output for the scene with multiple occlusions.



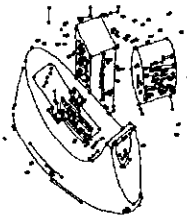
left image



right image



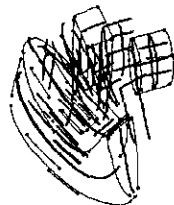
left edges



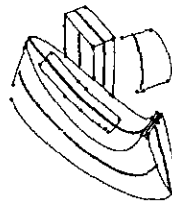
right edges



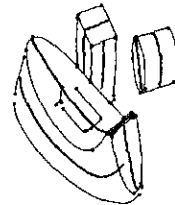
left symmetries



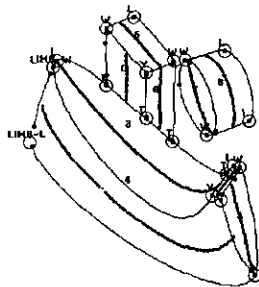
right symmetries



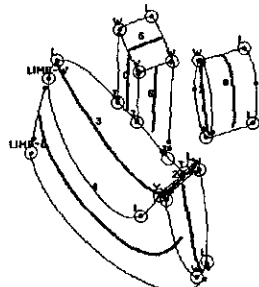
left selected ribbons



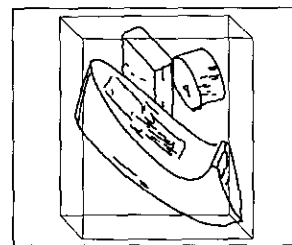
right selected ribbons



matched ribbons & junctions (left)



matched ribbons & junctions (right)



complete disparity output

FIG. 29. Results of a scene of curved objects.

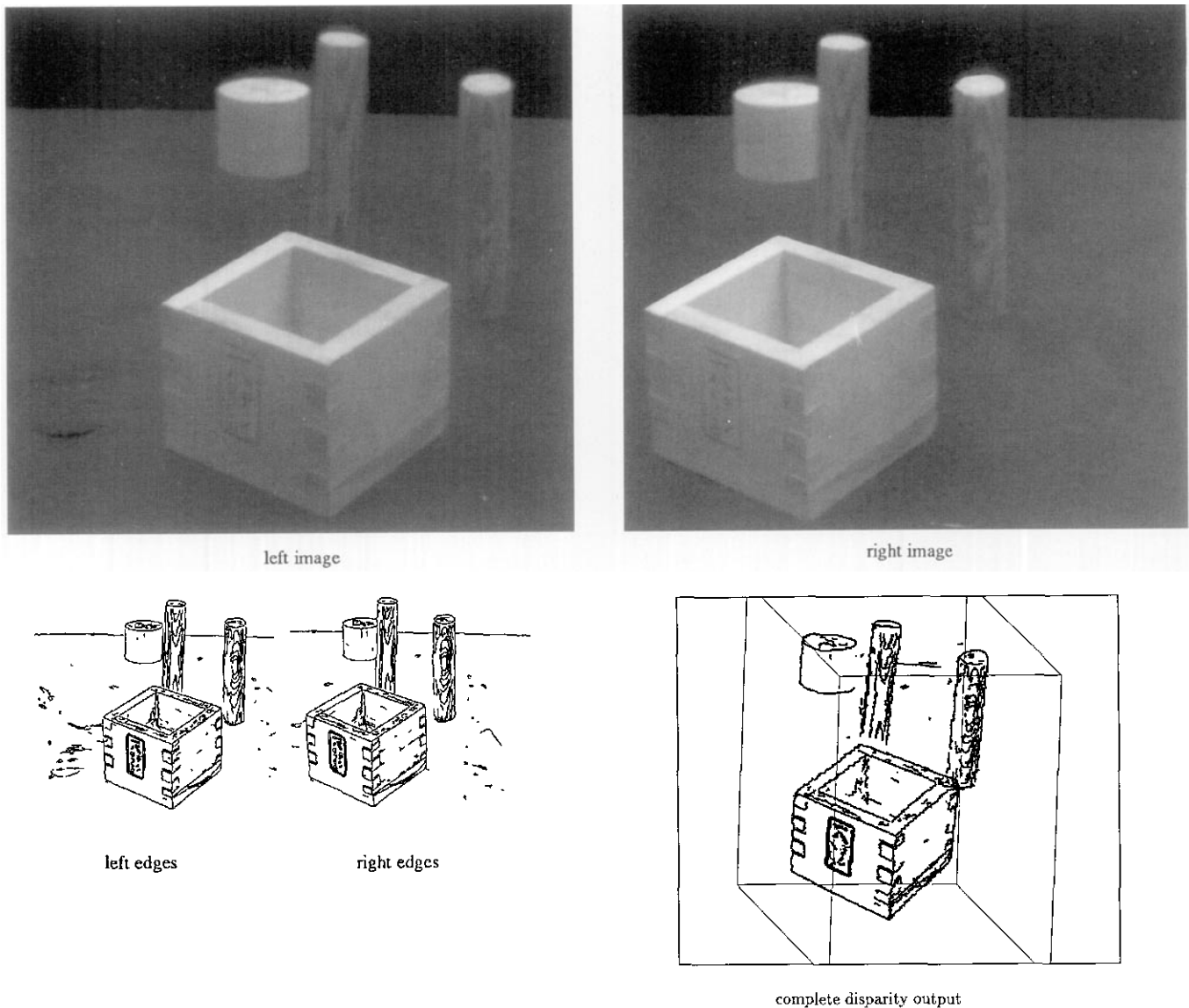


FIG. 30. Results of a scene of textured objects.

faces can be inferred from limb boundaries that are identified by our system.

We have not implemented surface interpolation procedures. However, we expect this to be a relatively easy task, as the surfaces have already been segmented into smooth regions.

Our system does require relatively high resolution images and probably would not be effective in highly textured and unstructured scenes where the monocular grouping processes fail (we can still match curves and get partial descriptions). To handle such scenes well, we would need to add yet another level of matching similar to what is used in area-based approaches to our system.

Note that the epipolar displacement of limb junctions

may be too small to be detectable due to the two low resolution of the images, and in such cases the limb junctions may be interpreted as edge-junctions. However, we argue that in such cases matching the limb edges directly, as if they were projected from orientation discontinuities, will also cause negligible error in the estimated disparity.

Our goal in stereo development has not been just to obtain a depth (or needle) map, but to compute abstract descriptions of the surfaces and objects visible in the scene. We believe that we have made major progress toward this goal, though much remains to be done. We attribute the success to the approach of combining descriptions and stereo matching, rather than viewing them as separate processes in a linear chain.

APPENDIX A

Proof of the T-Junction Behavior in Stereo Views

As shown in Fig. 5, let s be the arc length of the stem counting from the visible end to the occluded end, and s' be the arc length of the cap counting along an arbitrary direction. Also let $\mathbf{l}(s)$ and $\mathbf{l}'(s')$ be the position vectors of any point s on the stem and s' on the cap respectively with respect to the optical center of the camera at the original viewpoint.

For \bar{s} and \bar{s}' to be the contour points that project to a T junction, $\mathbf{l}(\bar{s}) \times \mathbf{l}'(\bar{s}') = 0$.

Let s_0 and s'_0 be the corresponding contour points that project to the T junction in the original view shown in Fig. 5. Then

$$\mathbf{l}(s_0) \times \mathbf{l}'(s'_0) = 0$$

Since s_0 is behind s'_0 along the line of sight, $[\mathbf{l}(s_0) - \mathbf{l}'(s'_0)]$ is in the same direction as $\mathbf{l}(s_0)$. It can also be observed that the sign of $[d\mathbf{l}(s)/ds|_{s=s_0} \times d\mathbf{l}'(s')/ds'|_{s'=s'_0}] \cdot \mathbf{l}'(s'_0)$ would depend on the chosen directions of s and s' , and in Fig. 5 it would be

$$\left[\frac{d\mathbf{l}(s)}{ds} \Big|_{s=s_0} \times \frac{d\mathbf{l}'(s')}{ds'} \Big|_{s'=s'_0} \right] \cdot \mathbf{l}'(s'_0) > 0.$$

Suppose now the viewpoint is moved slightly in an arbitrary direction \hat{t} by a magnitude Δt . Also let $s_1 = s_0 + \Delta s$ and $s'_1 = s'_0 + \Delta s'$ be the contour points on the stem and on the cap that project to the new T junction.

$$[\mathbf{l}(s_1) - (\Delta t)\hat{t}] \times [\mathbf{l}'(s'_1) - (\Delta t)\hat{t}] = 0$$

implies that

$$\begin{aligned} & \left[\mathbf{l}(s_0) + \frac{d\mathbf{l}(s)}{ds} \Big|_{s=s_0} \Delta s + \mathbf{O}_1((\Delta s)^2) - (\Delta t)\hat{t} \right] \\ & \times \left[\mathbf{l}'(s'_0) + \frac{d\mathbf{l}'(s')}{ds'} \Big|_{s'=s'_0} \Delta s' + \mathbf{O}_2((\Delta s')^2) - (\Delta t)\hat{t} \right] = 0 \end{aligned}$$

which in turn implies that

$$\begin{aligned} (\Delta s) \frac{d\mathbf{l}(s)}{ds} \Big|_{s=s_0} \times \mathbf{l}'(s'_0) + (\Delta s') \mathbf{l}(s_0) \times \frac{d\mathbf{l}'(s')}{ds'} \Big|_{s'=s'_0} \\ + (\Delta t)\hat{t} \times (\mathbf{l}(s_0) - \mathbf{l}'(s'_0)) + \mathbf{O}_3((\Delta s)^2) \\ + \mathbf{O}_4((\Delta s')^2) + \mathbf{O}_5(\Delta s \Delta s') + \mathbf{O}_6(\Delta t \Delta s) \\ + \mathbf{O}_7(\Delta t \Delta s') + \mathbf{O}_8(\Delta t^2) = 0 \end{aligned}$$

as $\mathbf{l}(s_0) \times \mathbf{l}'(s'_0) = 0$.

Dividing by Δt and taking the limit as $\Delta t \rightarrow 0$,

$$\begin{aligned} \left(\frac{ds}{dt} \Big|_{s=s_0} \right) \frac{d\mathbf{l}(s)}{ds} \Big|_{s=s_0} \times \mathbf{l}'(s'_0) + \left(\frac{ds'}{dt} \Big|_{s'=s'_0} \right) \mathbf{l}(s_0) \\ \times \frac{d\mathbf{l}'(s')}{ds'} \Big|_{s'=s'_0} + \hat{t} \times (\mathbf{l}(s_0) - \mathbf{l}'(s'_0)) = 0. \end{aligned}$$

Now if both sides of the equation are dot multiplied with $d\mathbf{l}'(s')/ds'|_{s'=s'_0}$, the term $ds'/dt|_{s'=s'_0}$ can be removed:

$$\begin{aligned} \left(\frac{ds}{dt} \Big|_{s=s_0} \right) \left[\frac{d\mathbf{l}(s)}{ds} \Big|_{s=s_0} \times \mathbf{l}'(s'_0) \right] \cdot \frac{d\mathbf{l}'(s')}{ds'} \Big|_{s'=s'_0} \\ + [\hat{t} \times (\mathbf{l}(s_0) - \mathbf{l}'(s'_0))] \cdot \frac{d\mathbf{l}'(s')}{ds'} \Big|_{s'=s'_0} = 0. \end{aligned}$$

Since $\mathbf{a} \cdot \mathbf{b} \times \mathbf{c} = \mathbf{a} \times \mathbf{b} \cdot \mathbf{c}$,

$$\begin{aligned} \left(\frac{ds}{dt} \Big|_{s=s_0} \right) \left[\frac{d\mathbf{l}'(s')}{ds'} \Big|_{s'=s'_0} \times \frac{d\mathbf{l}(s)}{ds} \Big|_{s=s_0} \right] \cdot \mathbf{l}'(s'_0) \\ + \hat{t} \cdot \left[(\mathbf{l}(s_0) - \mathbf{l}'(s'_0)) \times \frac{d\mathbf{l}'(s')}{ds} \Big|_{s'=s'_0} \right] = 0 \end{aligned}$$

which implies that

$$\frac{ds}{dt} \Big|_{s=s_0} = \frac{\hat{t} \cdot \left[(\mathbf{l}(s_0) - \mathbf{l}'(s'_0)) \times \frac{d\mathbf{l}'(s')}{ds'} \Big|_{s'=s'_0} \right]}{\left[\frac{d\mathbf{l}(s)}{ds} \Big|_{s=s_0} \times \frac{d\mathbf{l}'(s')}{ds'} \Big|_{s'=s'_0} \right] \cdot \mathbf{l}'(s'_0)}.$$

Since the sign of

$$\left[\frac{d\mathbf{l}(s)}{ds} \Big|_{s=s_0} \times \frac{d\mathbf{l}'(s')}{ds'} \Big|_{s'=s'_0} \right] \cdot \mathbf{l}'(s'_0)$$

is fixed (positive in the example here), and $(\mathbf{l}(s_0) - \mathbf{l}'(s'_0))$ has the same direction as $\mathbf{l}(s_0)$, the sign of $ds/dt|_{s=s_0}$ only depends upon the sign of the dot product

$$\hat{t} \cdot \left[\mathbf{l}(s_0) \times \frac{d\mathbf{l}'(s')}{ds'} \Big|_{s'=s'_0} \right].$$

Let Π be the plane perpendicular to $\mathbf{l}(s_0) \times d\mathbf{l}'(s')/ds'|_{s'=s'_0}$ and containing $\mathbf{l}(s_0)$, which is actually the plane formed by the line of sight of s'_0 (same as that of s_0) and

the tangent to the cap contour (occluding contour) at s'_0 . Unless the image plane moves along plane Π , i.e., it moves exactly toward or away from the contour point s_0 ; otherwise ds would not be zero. That means if the viewpoint is moved sideways in any direction, $s_1 \neq s_0$ and the point projecting to the T junction does move along the stem contour.

Plane Π divides the 3-D space into two half-spaces, one with more of the occluded surface patch S visible, and the other less. If \hat{t} is not pointing along plane Π but to the half-space in which more of the occluded surface is visible (the viewpoint moves from a less-exposure view to a more-exposure view), $ds > 0$ and therefore the stem contour point projecting to the T junction moves toward the occluded portion of the stem. In our example here the T junction will appear to move up across the epipolar lines in the image as the projection of the stem's occluded portion goes up across the epipolar lines. If \hat{t} is pointing to the other half-space (the viewpoint moves from a more-exposure view to a less-exposure view), $ds < 0$ and therefore the stem contour point projecting to the T junction moves away from the occluded portion of the stem. In our example here the T junction will appear to move down across the epipolar lines in the image. This proves the constraint.

APPENDIX B

Proof of the Limb-Junction Behavior in Stereo Views

As shown in Fig. 8, let s be the arc length of the terminator contour counting from the front edge to the back edge. Also let $\mathbf{l}(s)$ be the position vector of any point s on the terminator contour with respect to the optical center of the camera at the original viewpoint, and $\hat{r}(s)$ and $\hat{n}(s)$ be the surface normals of the limb surface L and the terminator surface T respectively at the point s on the terminator contour.

For \bar{s} to be the terminator contour point projecting to a limb junction, $\mathbf{l}(\bar{s}) \cdot \hat{r}(\bar{s}) = 0$.

Let s_0 be the terminator contour point that projects to the limb junction. Then

$$\mathbf{l}(s_0) \cdot \hat{r}(s_0) = 0.$$

Suppose now the viewpoint is moved slightly in an arbitrary direction \hat{t} by a magnitude Δt . Also let $s_1 = s_0 + \Delta s$ be the terminator contour point projecting to the new limb junction.

$$[\mathbf{l}(s_1) - (\Delta t)\hat{t}] \cdot \hat{r}(s_1) = 0$$

implies that

$$\left[\mathbf{l}(s_0) + \frac{d\mathbf{l}(s)}{ds} \Big|_{s=s_0} \Delta s + \mathbf{O}_1((\Delta s)^2) - (\Delta t)\hat{t} \right] \cdot \left[\hat{r}(s_0) + \frac{d\hat{r}(s)}{ds} \Big|_{s=s_0} \Delta s + \mathbf{O}_2((\Delta s)^2) \right] = 0$$

which in turn implies that

$$\begin{aligned} (\Delta s) \frac{d\hat{r}(s)}{ds} \Big|_{s=s_0} \cdot \mathbf{l}(s_0) - (\Delta t)\hat{t} \cdot \hat{r}(s_0) + O_3((\Delta s)^2) \\ + O_4(\Delta s \Delta t) = 0 \end{aligned}$$

as $\mathbf{l}(s_0) \cdot \hat{r}(s_0) = 0$ and $d\mathbf{l}(s)/ds|_{s=s_0} \perp \hat{r}(s_0)$.

Dividing by Δt and taking the limit as $\Delta t \rightarrow 0$,

$$\frac{ds}{dt} \Big|_{s=s_0} = \frac{\hat{t} \cdot \hat{r}(s_0)}{\frac{d\hat{r}(s)}{ds} \Big|_{s=s_0} \cdot \mathbf{l}(s_0)}$$

Since $d\hat{r}(s)/ds|_{s=s_0}$ is in the same direction as $\mathbf{l}(s_0)$ and thus $d\hat{r}(s)/ds|_{s=s_0} \cdot \mathbf{l}(s_0) > 0$, the sign of $ds/dt|_{s=s_0}$ only depends upon the sign of the dot product

$$\hat{t} \cdot \hat{r}(s_0)$$

Let Π be the plane perpendicular to $\hat{r}(s_0)$ and containing $\mathbf{l}(s_0)$, which is actually the tangential plane to the limb surface at the point s_0 . Unless the image plane moves along plane Π , i.e., it moves exactly towards or away from the terminator contour point s_0 , otherwise ds would not be zero. That means if the viewpoint is moved sideways in any direction, $s_1 \neq s_0$ and the point projecting to the limb junction does move along the terminator contour.

Plane Π divides the 3-D space into two half-spaces, one with more of the occluded portion of the limb surface L visible, and the other not. If \hat{t} is not pointing along plane Π but to the half-space in which more of the occluded portion of the limb surface is visible (the viewpoint moves from a less-exposure view to a more-exposure view), $ds > 0$ and therefore the terminator contour point projecting to the limb junction moves toward the back edge. In our example here the limb junction will appear to move up across the epipolar lines in the image as the projection of the back edge goes up across the epipolar lines. If \hat{t} is pointing to the other half-space (the viewpoint moves from a more-exposure view to a less-exposure view), $ds < 0$ and therefore the terminator contour point projecting to the limb junction moves away from the back edge. In our example here the limb junction will appear to move down

across the epipolar lines in the image. This proves the constraint.

REFERENCES

1. N. Ayache and B. Faverjon, Fast stereo matching of edge segments using prediction and verification of hypotheses, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, June 19–23, 1985*, pp. 662–664.
2. H. H. Baker, T. O. Binford, J. Malik, and J. Meller, Progress in stereo mapping, in *Proceedings of the DARPA Image Understanding Workshop, Arlington, VA, June 23, 1983*, pp. 327–335.
3. P. N. Belhumeur and D. Mumford, A Bayesian treatment of the stereo correspondence problem using half-occluded regions, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Champaign, IL, June 1992*, pp. 506–512.
4. J. F. Canny, A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(6), 1986, 679–698.
5. R. Chung and R. Nevatia, Use of monocular groupings and occlusion analysis in a hierarchical stereo system, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Maui, HI, June 1991*, pp. 50–56.
6. S. D. Cochran and G. Medioni, Accurate surface description from binocular stereo, in *Proceedings of the Workshop in Interpretation of 3D Scenes, Austin, TX, Nov. 27–29, 1989*, pp. 16–23.
7. U. R. Dhond and J. K. Aggarwal, Structure from stereo—A review, *IEEE Trans. Syst. Man Cybernet.* **19**(6), 1989, 1489–1510.
8. U. R. Dhond and J. K. Aggarwal, Computing stereo correspondences in the presence of narrow occluding objects, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Champaign, IL, June 1992*, pp. 758–760.
9. M. Drumheller and T. Poggio, On parallel stereo, in *Proceedings of the IEEE Conference on Robotics and Automation, San Francisco, CA, Apr. 1986*, pp. 1439–1448.
10. S. Ganapathy, Reconstruction of scenes containing polyhedra from stereo pair of views, Ph.D. thesis, Stanford University, Stanford, CA, 1976.
11. D. Geiger, B. Ladendorf, and A. Yuille, Occlusions and binocular stereo, in *Proceedings of the European Conference on Computer Vision, Santa Margherita Ligure, Italy, May 1992*.
12. W. E. L. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Visual System*, MIT Press, Cambridge, MA, 1981.
13. M. Herman and T. Kanade, The 3D MOSAIC scene understanding system: Incremental reconstruction of 3D scenes from complex images, Technical Report CMU-CS-84-102, Carnegie-Mellon University, Pittsburgh, PA, Feb. 1984.
14. W. Hoff and N. Ahuja, Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(2), 1989, 121–136.
15. J. J. Hopfield, Neurons with graded response have collective computational properties like those of two-state neurons, *Proc. Natl. Acad. Sci. USA* **81**, May 1984, 3088–3092.
16. J. J. Hopfield and D. W. Tank, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, Apr. 1982, 2554–2558.
17. S. Liebes, Geometric constraints for interpreting images of common structural elements: Orthogonal trihedral vertices, in *Proceedings of the DARPA Image Understanding Workshop, Apr. 1981*.
18. H. S. Lim and T. O. Binford, Structural correspondence in stereo vision, in *Proceedings of the DARPA Image Understanding Workshop, Los Angeles, CA, 1987*, pp. 794–808.
19. H. S. Lim and T. O. Binford, Curved surface reconstruction using stereo correspondence, in *Proceedings of the DARPA Image Understanding Workshop, Cambridge, MA, Apr. 1988*, pp. 809–819.
20. J. J. Little and W. E. Gillett, Direct evidence for occlusion in stereo and motion, *Image Vision Comput.* **8**(4), 1990, 328–340.
21. D. Marr and T. Poggio, Cooperative computation of stereo disparity, *AAAS Sci.* **194**(4262), 1976, 283–287.
22. D. Marr and T. Poggio, A theory of human stereo vision, Technical Report AI Memo 451, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Cambridge, MA, Nov. 1977.
23. G. Medioni and R. Nevatia, Segment-based stereo matching, *Comput. Graphics Image Process.* **31**(1), 1985, 2–18.
24. R. Mohan and R. Nevatia, Using perceptual organization to extract 3-D structures, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(11), 1989, 1121–1139.
25. R. Mohan and R. Nevatia, Perceptual organization for scene segmentation and description, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(6), 1992, 616–635.
26. V. Nalwa, Line-drawing interpretation: A mathematical framework, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1988*, pp. 18–31.
27. K. Prazdny, Detection of binocular disparities, *Biol. Cybernet.* **52**, 1985, 93–99.
28. P. Saint-Marc and G. Medioni, Adaptive smoothing for feature extraction, in *Proceedings of the DARPA Image Understanding Workshop, Boston, MA, Apr. 1988*, pp. 1100–1113.
29. D. Terzopoulos, Regularization of inverse visual problems involving discontinuities, *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 1986, 413–424.